Econometric Theory, **30**, 2014, 839–881.
doi:10.1017/S0266466613000492

PANEL STRUCTURAL MODELING WITH WEAK INSTRUMENTATION AND COVARIANCE RESTRICTIONS

JOHN C. CHAO
University of Maryland

This paper considers estimating a panel data simultaneous equations model under both coefficient and covariance matrix restrictions in a scenario where one or the other set of identifying restrictions may be invalid or may hold only weakly. We study the limiting properties of various estimators in an asymptotic framework, which takes both the cross-sectional dimension N and the time dimension T to infinity. In this setting as in the pure cross-sectional setup, the performance of the 2SLS estimator depends on the strength of the identifying conditions imposed on the coefficients of the model, and it fails to be consistent once these conditions break down sufficiently resulting in instruments that are too weakly correlated with the endogenous regressors. On the other hand, the between-group (BG) estimator is consistent and asymptotically normal even when coefficient restrictions fail, but it has the shortcoming that its precision depends only on variations in the cross-sectional dimension; and, hence, it is less efficient and has slower rate of convergence than alternatives, which make better use of the large time dimension. A GMM estimator, which combines the moment conditions of the BG estimator with that of the within-group IV estimator, is more robust to instrument weakness than 2SLS and is more efficient than the BG estimator, but it has a second-order bias even under strong instruments if the assumed covariance restrictions do not hold. To remedy the deficiency of the aforementioned estimators, we propose in this paper two new model averaging estimators, which are weighted averages of the GMM estimator and a bias-corrected GMM estimator. The two proposed estimators have weighting functions that depend on alternative transformations of the Bayesian Information Criterion (BIC), which is employed here to assess the validity of the covariance restrictions. We show that these new estimators have some nice robustness properties against possible failure of either the coefficient restrictions or the covariance restrictions.

JEL classification: C13, C33.

The author would like to thank the Guest Editor, Bruce Hansen, and two anonymous referees for their very helpful comments and suggestions. With great affection, this paper is dedicated to my mentor and friend, Peter C. B. Phillips. Address correspondence to John C. Chao, Department of Economics, University of Maryland, College Park, MD 20742. e-mail: chao@econ.umd.edu.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

1. INTRODUCTION

Parameters in linear simultaneous equations models have traditionally been identified by imposing certain conditions on the coefficients of the model. The typical identification strategy requires the presence of instrumental variables, which are excluded from the structural equation of interest but which must also appear elsewhere in the system as significant explanatory variables for the key endogenous regressors whose causal effects one is trying to estimate. In situations where the assumed conditions on the coefficient parameters do not hold and there are no other identifying restrictions, both the commonly used instrumental variables (IV) estimator and the associated test statistics would be badly behaved. The precise nature of this bad behavior has been elegantly studied in a pioneering paper by Phillips (1989).

Putting restrictions on the coefficients of the simultaneous equations model is, of course, not the only method for achieving identification. Another possibility for obtaining identification is to place restrictions on the error covariance matrix, as have been studied by Hausman and Taylor (1983) and Hausman, Newey, and Taylor (1987). Given that it is often difficult to specify good-quality instruments, it would seem that applied researchers would have incentives to explore various alternative avenues for achieving identification. Moreover, in a given empirical situation, there might be uncertainty about the validity of one set of identifying restrictions vis-à-vis another, so that estimators, which have some robustness properties against the partial failure of some of the restrictions might be desired.

The purpose of the present paper is to study these issues in the context of a panel data simultaneous equations (or IV regression) model. More specifically, we consider estimating this model under both coefficient and covariance matrix restrictions but in a situation where one or the other set of restrictions may not be valid or may hold only weakly. We propose several estimators, which can take advantage of the possibility that there may be restrictions on the error component covariance matrix of this model. First, we consider a generalized method of moments (GMM) estimator which combines the moment conditions of the between-group regression with that of the within-group regression. Under restrictions on the random effects covariance matrix, this estimator will be consistent and asymptotically normal even in situations where the instruments of our simultaneous equations model do not have sufficient correlation with the endogenous regressors to ensure consistent estimation using the two-stage least squares (2SLS) estimator. However, a shortcoming of the GMM estimator is that when the assumed covariance restrictions do not hold, then it is not only inconsistent when instruments are sufficiently weak but also suffers from a second-order bias even under strong instruments. To remedy this deficiency, we propose in this paper two additional "model averaging" estimators, which are weighted averages of the GMM estimator and a bias-corrected GMM estimator. These two new estimators have weighting functions that depend on alternative transformations of the Bayesian Information Criterion (BIC), which is employed here to assess the

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

validity of the covariance restrictions. An interesting property of both of these estimators is that they have some nice adaptive capabilities in the sense that both estimators are asymptotically equivalent to the GMM estimator when the covariance restrictions are valid, but they correct for the bias of GMM when these restrictions fail.

There have been other papers, which have studied the finite sample properties of IV/GMM estimators in the context of panel data. In particular, Alonso-Borrego and Arellano (1999) have introduced symmetrically normalized GMM estimators for estimating the autoregressive parameter of a panel data AR(1) model and have shown via a simulation study that these estimators have smaller median bias than the ordinary GMM estimator. Bun and Windmeijer (2007) have also studied the panel AR(1) model, focusing, in particular, on the weakly identified case where the AR(1) process is persistent and where the variance of the unobserved random effect is of the same magnitude as the variance of the idiosyncratic disturbances. They present, among other things, an interesting analysis to explain why the system GMM estimator in this case has a lower bias than a GMM procedure, which only makes use of moment conditions obtained from representing the model in first-differences. Our paper complements these earlier works. More specifically, our study differs from these papers in that we analyze a panel data simultaneous equations model in the tradition of Baltagi (1981), Prucha (1985), and Baltagi and Li (1992), but unlike these latter authors who only consider the case of strong identification by coefficient restrictions, we explicitly consider the possibility of weak instruments and the possible presence of restrictions on the covariation of individual effects. In addition, none of these papers has considered combining estimators to obtain a procedure which may be consistent under alternative identifying restrictions.

The research reported in this paper has been greatly influenced by the work of Peter Phillips, and this seems necessarily so given the breadth and depth of what he has contributed in this area. Thus, we will begin in Section 2 with a brief review of some of his papers, which are most closely related to the subject matter studied here. We then proceed to Section 3 to lay out the model and assumptions considered in this paper. Section 4 discusses estimation in situations where there are covariance restrictions but where the available instruments might be of poor quality. Section 5 analyzes the case where the assumed covariance restrictions do not hold and also introduces as well as studies the asymptotic properties of our new model averaging estimators. A small Monte Carlo study, which evaluates the finite sample properties of a variety of alternative estimators, is given in Section 6. Finally, some concluding remarks are offered in Section 7, while proofs of the main results are collected in the Appendix. A supplement to this paper is also available online, and it contains additional technical details, including additional lemmas as well as technical arguments and calculations used to establish some of the results of this paper. Interested readers may refer to the Supplementary Material associated with this article, available at Cambridge Journals Online (journals.cambridge.org/ect).

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

2. PHILLIPS' CONTRIBUTIONS TO IV/GMM MODELING UNDER PARTIAL AND WEAK IDENTIFICATION

Peter Phillips has made a number of fundamental contributions to the analysis of IV/GMM models under conditions of either near or complete identification failure. While some previous authors such as Liu (1960) and Sims (1980) have expressed concerns about the credibility of exclusion restrictions imposed in empirical work for the sake of trying to achieve identification, a rigorous analysis of possible consequences when such restrictions failed to hold was not available until the pathbreaking work of Phillips (1989). In particular, Phillips (1989) showed that, for a totally unidentified simultaneous equations model, the (unstandardized) IV/2SLS estimator was inconsistent and, in fact, converged weakly to a mixed-normal distribution, which carried no information at all about the structural parameter of interest, reflecting the lack of identification. In addition, an interesting connection between exact finite sample distribution of IV estimators and their asymptotic distribution was pointed out by this paper and by the results of Phillips' earlier work (Phillips 1980, 1984, 1985), which collectively showed that, with identification weakness, the exact sampling distribution derived under the explicit assumption of normality takes on the added significance of also being the correct limiting distribution under more general error processes. Indeed, by combining the noncentral exact distributional results presented in Phillips (1980) with the partial identification limit theory given in Phillips (1989), one could obtain a general limit theory for IV estimation in linear simultaneous equations model with weak instruments. Hence, the partial identification results obtained by Phillips in this series of papers anticipated subsequent developments in the weak instrument literature, as it bore a close relationship to the local-to-zero weak instrument asymptotics of Staiger and Stock (1997), which also resulted in IV estimators having asymptotic distribution with the same mixed-normal structure.

The analysis of Phillips (1989) was later extended in a number of interesting directions by Peter Phillips himself, his co-authors, and others. In particular, Choi and Phillips (1992) generalized the analysis of Phillips (1989) to the case of a partially identified, linear IV regression model where certain linear combinations of the structural parameter may be estimable. The case of a simultaneous equations system with an identity was taken up by Phillips (2006), which shows that, interestingly enough, the limiting distribution of the IV estimator in that case is inverse Gaussian when instruments are very weak. A very general and deep analysis of the problem of weak identification was given in Han and Phillips (2006) where they consider possibly nonlinear GMM models, which embody both the possibility of weak identification with a fixed number of moment conditions, as in Stock and Wright (2000), and the many weak instrument setup of Chao and Swanson (2005) and Stock and Yogo (2005). An important technical contribution of this paper is the development of a general framework for using epiconvergence methods to prove limit theorems for possibly weakly identified models. This methodology is likely to be very useful to future researchers working in this area.

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Taken as a whole, Phillips' research has done a great deal to enhance and deepen our understanding of the problems that can occur when applying conventional estimators and test procedures in the presence of identification failure or identification weakness. His analysis has also yielded substantial insights on how to design reliable statistical procedures in such situations. Hence, it is very much in the tradition of his work that the present paper is written. We will show below that when a panel data IV regression model is only weakly identified by coefficient restrictions, the 2SLS estimator will be inconsistent and have a mixed-normal limiting distribution analogous to that derived by Phillips (1989) in the cross-sectional context. However, for the problem we study here, we will be able to construct alternative estimators which will be consistent and can be used to construct valid test procedures, provided that there are additional covariance restrictions.

3. MODEL AND ASSUMPTIONS

Consider the following error component simultaneous equations model (ECSEM)

$$y_{1,it} = Y_{2,it}'\beta^0 + \mu_i + u_{it}, \tag{1}$$

$$Y_{2,it} = \Pi' Z_{it} + \eta_i + v_{it}, \tag{2}$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Here, $y_{1,it}$ (1×1) and $Y_{2,it}$ $(G \times 1)$ are the endogenous variables of the system, Z_{it} $(K \times 1)$ are the exogenous variables (or instruments), and μ_i (1×1) and η_i $(G \times 1)$ represent random effects, which appear in the structural equation (1) and the first-stage equation (2), respectively. Throughout this paper, we shall consider only the case where Z_{it} is nonrandom, although our conditions and results (given below) could be readily modified to accommodate the case with random (exogenous) instruments. We also do not consider the case where there are included exogenous variables in the structural equation (1) in order to lessen the amount of notational burden. However, it is easily seen that our results could be extended in a straightforward manner to the case where there are included exogenous variables.

Next, define $y_{1,i\cdot} = (y_{1,i1,\ldots,}y_{1,iT})'$, $Y_{2,i\cdot} = (Y_{2,i1,\ldots,}Y_{2,iT})'$, $Z_{i\cdot} = (Z_{i1,\ldots,}Z_{iT})'$, $u_{i\cdot} = (u_{i1,\ldots,}u_{iT})'$, and $V_{i\cdot} = (v_{i1,\ldots,}v_{iT})'$ for $i = 1, \ldots, N$; and let $y_1 = (y_{1,1\cdot,\ldots,}\, y_{1,N\cdot})'$, $Y_2 = (Y_{2,1\cdot,\ldots,}\, Y_{2,N\cdot})'$, $Z = (Z_{1\cdot,\ldots,}\, Z_{N\cdot})'$, $u = (u_{1\cdot,\ldots,}\, u_{N\cdot})'$, and $V = (V_{1\cdot,\ldots,}\, V_{N\cdot})'$. Also, define $\mu = (\mu_1, \ldots, \mu_N)'$ and $\eta = (\eta_1, \ldots, \eta_N)'$ and let ι_T be a $T \times 1$ vector of ones. If follows that with these notations, we can stack the NT observations and write (1) and (2) as

$$y_1 = Y_2\beta^0 + (I_N \otimes \iota_T)\,\mu + u, \tag{3}$$

$$Y_2 = Z\Pi + (I_N \otimes \iota_T)\,\eta + V. \tag{4}$$

To facilitate the statement of our assumptions, we introduce further notations. In the sequel, $tr(\cdot)$ denotes the trace of a matrix, " > 0" denotes positive definiteness when applied to matrices, C is a generic constant that may be different

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

in different usage, and $\|\cdot\|$ denotes the Frobenius or Euclidean norm. In addition, $P_X = X(X'X)^{-1}X'$ denotes the matrix that projects orthogonally onto the range space of X and $Q_X = I - P_X$. Furthermore, let $\xi_{it} = \left(u_{it}, v_{it}'\right)'$ and let $\xi_{a,it}$ denote the a^{th} element of ξ_{it}, and we can define the fourth order cumulant of $\left(\xi_{a,it}, \xi_{b,i(t+\ell)}, \xi_{c,i(t+m)}, \xi_{d,i(t+n)}\right)$ as follows

$$\kappa^{(i)}_{abcd}(t.t+\ell, t+m.t+n) = E\left[\xi_{a,it}\xi_{b,i(t+\ell)}\xi_{c,i(t+m)}\xi_{d,i(t+n)}\right] - E\left[\widetilde{\xi}_{a,it}\widetilde{\xi}_{b,i(t+\ell)}\widetilde{\xi}_{c,i(t+m)}\widetilde{\xi}_{d,i(t+n)}\right],$$

where $\left\{\widetilde{\xi}_{it}\right\}$ is a Gaussian sequence whose first two moments match that of $\left\{\xi_{it}\right\}$. By analogous notation, we also define

$$\kappa^{(i)}_{u}(t.t+\ell, t+m.t+n) = E\left[u_{it}u_{i(t+\ell)}u_{i(t+m)}u_{i(t+n)}\right] - E\left[\widetilde{u}_{it}\widetilde{u}_{i(t+\ell)}\widetilde{u}_{i(t+m)}\widetilde{u}_{i(t+n)}\right]$$

$$\kappa^{(i)}_{v,abcd}(t.t+\ell, t+m.t+n) = E\left[v_{a,it}v_{b,i(t+\ell)}v_{c,i(t+m)}v_{d,i(t+n)}\right] - E\left[\widetilde{v}_{a,it}\widetilde{v}_{b,i(t+\ell)}\widetilde{v}_{c,i(t+m)}\widetilde{v}_{d,i(t+n)}\right].$$

In addition, let

$$\Sigma_{u,NT} = \begin{pmatrix} \Sigma^{(1)}_{u,T} & 0 & \cdots & 0 \\ 0 & \Sigma^{(2)}_{u,T} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \Sigma^{(N)}_{u,T} \end{pmatrix}, \tag{5}$$

where

$$\Sigma^{(i)}_{u,T} = \begin{pmatrix} \gamma^{(i)}_u(1,1) & \gamma^{(i)}_u(1,2) & \gamma^{(i)}_u(1,3) & \cdots & & \cdots & \gamma^{(i)}_u(1,T) \\ \gamma^{(i)}_u(2,1) & \gamma^{(i)}_u(2,2) & \gamma^{(i)}_u(2,3) & \ddots & & & \vdots \\ \gamma^{(i)}_u(3,1) & \gamma^{(i)}_u(3,2) & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \gamma^{(i)}_u(T-1,T-1) & & \gamma^{(i)}_u(T-1,T) \\ \gamma^{(i)}_u(T,1) & \cdots & \cdots & \cdots & \gamma^{(i)}_u(T,T-1) & & \gamma^{(i)}_u(T,T) \end{pmatrix}, \tag{6}$$

with $\gamma^{(i)}_u(s,t) = E\left[u_{is}u_{it}\right]$ for $i = 1, \ldots N$; $s = 1, \ldots, T$; and $t = 1, \ldots, T$. We make the following assumptions.

Assumption 1 (Local-to-zero parameterization).

$\Pi = \Pi_{NT} = N^{-\delta_1}T^{-\delta_2}D^0$ for $\delta_1 \geq 0$ and $\delta_2 \geq 0$,

where D^0 is a $K \times G$ nonrandom matrix of full column rank $G \leq K$.

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Assumption 2 (Exogenous variables).

(a) There exists constant C such that $\sup_{i\geq 1,t\geq 1}\left\|Z_{it}\right\| \leq C < \infty$.

(b) There exist positive definite matrices M_P and M_Q such that

$$\frac{Z'\left(I_N \otimes P_{\iota_T}\right)Z}{NT} \to M_P,$$
$$\frac{Z'\left(I_N \otimes Q_{\iota_T}\right)Z}{NT} \to M_Q,$$

as $N, T \to \infty$ (jointly) with $N/T \to c$ for some constant c with $0 < c < \infty$.

(c) There exists a nonrandom vector m_Z such that $N^{-1}\sum_{i=1}^N \overline{Z}_i = N^{-1}\sum_{i=1}^N \left(T^{-1}\sum_{t=1}^T Z_{it}\right) \to m_Z$ as $N, T \to \infty$ (jointly) with $N/T \to c$ for some constant c with $0 < c < \infty$.

Assumption 3. $\left\{\xi_{it}\right\}$ is a double-indexed sequence of random vectors, which are independent across the index i. We assume that $E\left[\xi_{it}\right] = 0$ for all i and t. Moreover, there exists a constant C such that

$$\sum_{\ell=1}^{\infty} \sup_{i\geq 1,t\geq 1}\left\|E\xi_{it}\xi'_{i(t+\ell)}\right\| \leq C < \infty,$$

$$\sum_{\ell=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \sup_{i\geq 1,t\geq 1}\left|\kappa^{(i)}_{abcd}(t.t+\ell, t+m.t+n)\right| \leq C < \infty \ \forall a,b,c,d \leq G+1.$$

Assumption 4 (Random effects).

(a) $\left\{\eta_i\right\} \equiv i.i.d.\left(0, \Sigma_{\eta\eta}\right)$, where $\Sigma_{\eta\eta} > 0$; there exists a constant C such that $E\left[\left\|\eta_i\right\|^8\right] \leq C < \infty$; and η_i is independent of ξ_{jt} for all i, j, and t.

(b) $\left\{\mu_i\right\} \equiv i.i.d.\left(0, \sigma^2_\mu\right)$, where $\sigma^2_\mu > 0$; there exists a constant C such that $E\left[\mu_i^8\right] \leq C < \infty$, and μ_i is independent of ξ_{jt} for all i, j, and t.

Assumption 5 (Covariance restriction). $\sigma_{\eta\mu} = E\left[\eta_i\mu_i\right] = 0$.

Assumption 6. There exists a positive constant C such that for all N, T sufficiently large

$$\lambda_{\min}\left(\Sigma_{u,NT}\right) \geq C > 0, \text{ and } \lambda_{\min}\left(\Psi_{NT}\right) \geq C > 0,$$

where

$$\Psi_{NT} = \begin{pmatrix} \sigma^2_\mu N^{-1}\sum_{i=1}^N \overline{Z}_i\overline{Z}'_i & 0 & N^{-1}\sum_{i=1}^N \overline{Z}_i E\left[\mu_i^2\eta'_i\right] \\ 0 & N^{-1}\sum_{i=1}^N \left(\Sigma_{\eta\eta}\otimes \overline{Z}_i\overline{Z}'_i\right) & N^{-1}\sum_{i=1}^N \left(I_G \otimes \overline{Z}_i\right) E\left[\mu_i\eta_i\eta'_i\right] \\ N^{-1}\sum_{i=1}^N \left(E\left[\mu_i^2\eta'_i\right]\right)'\overline{Z}'_i & N^{-1}\sum_{i=1}^N E\left[\mu_i\eta_i\eta'_i\right]\left(I_G\otimes \overline{Z}'_i\right) & E\left[\mu_i^2\eta_i\eta'_i\right] \end{pmatrix}$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Remark 3.1.

(i) Assumption 1 allows for the possibility of weak instruments by specifying a general local-to-zero structure for our panel IV regression model. The weak-instrument setup we consider here, thus, extends the local-to-zero framework of Staiger and Stock (1997) to a panel data setting where both the cross-sectional and the time dimensions are large.

(ii) Note that Assumption 3 allows for heterogeneity across both i and t.

(iii) Assumption 3 is very similar to Assumption A* given in Andrews (1991), except that here, because we are dealing with panel data, we also take the supremum over the cross sectional units i. Similar cumulant conditions for the fourth-order stationary case can also be found in Hannan (1970) and Anderson (1971).

4. ESTIMATION UNDER COVARIANCE RESTRICTIONS

In this section, we consider the problem of estimating the structural parameter (vector) β in the case where the instruments provided by coefficient restrictions might be weak but where additional identifying restrictions are available in light of Assumption 5. A commonly used estimator of β in the present limited information context is the two-stage least squares (2*SLS* henceforth) estimator

$$\widehat{\beta}_{2SLS} = \left(Y_2' P_Z Y_2\right)^{-1} Y_2' P_Z y_1, \tag{7}$$

or its generalized least squares (GLS) analog (cf. Hsiao, 2003, Section 5.3.1). However, as the following theorem shows, the 2*SLS* estimator does not make use of the additional covariance restrictions and, thus, is not consistent if the instruments provided by coefficient restrictions are sufficiently weak.

THEOREM 4.1 (2*SLS*).

Suppose that Assumptions 1–6 hold and suppose that $N, T \to \infty$ *such that* $N/T \to c$ *for some constant* c *with* $0 < c < \infty$. *Then,*

(a) *if* $\delta_1 + \delta_2 > 1/2$,

$$\widehat{\beta}_{2SLS} - \beta^0 \Longrightarrow \int_{\mathbb{R}^{KG}} N\left(0,\ \sigma_\mu^2 S(X)\right) pdf(X) dX,$$

where

$$S(X) = \left(X' M X\right)^{-1} X' M_P X \left(X' M X\right)^{-1},$$
$$vec(X) \sim N\left(0, \Sigma_{\eta\eta} \otimes M^{-1} M_P M^{-1}\right), \tag{8}$$

and where $M = M_P + M_Q$.

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

(b) *if* $\delta_1 + \delta_2 = 1/2$,

$$\widehat{\beta}_{2SLS} - \beta^0 \Longrightarrow \int_{\mathbb{R}^{KG}} N\left(0,\ \sigma_\mu^2 S(A(X))\right) pdf(X)dX,$$

where $S(\cdot)$ *is as defined in (8) above, so that*

$$S(A(X)) = \left[A(X)'MA(X)\right]^{-1} A(X)'M_P A(X) \left[A(X)'MA(X)\right]^{-1}$$

and

$$A(X) = c^{1/2-\delta_1} D^0 + X,$$

where X *is a normally distributed random matrix as defined in (8) above.*

(c) *if* $0 \leq \delta_1 + \delta_2 < 1/2$,

$$T^{1/2-\delta_1-\delta_2}\left(\widehat{\beta}_{2SLS} - \beta^0\right) \Longrightarrow N\left(0, \sigma_\mu^2 c^{2\delta_1-1}\left[D^{0\prime}MD^0\right]^{-1} D^{0\prime}M_P D^0 \left[D^{0\prime}MD^0\right]^{-1}\right).$$

Another estimator, which turns out to have some interesting asymptotic properties in the current setup, is the between-group (OLS) estimator

$$\widehat{\beta}_{BG} = \left(Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2\right)^{-1} Y_2'\left[I_N \otimes P_{\iota_T}\right] y_1. \tag{9}$$

This estimator is consistent and asymptotically normal under the covariance restrictions given by Assumption 5, as shown in the theorem given below.

THEOREM 4.2 (*BG*).

Suppose that Assumptions 1–5 hold and suppose that $N, T \to \infty$ *such that* $N/T \to c$ *for some constant* c *with* $0 < c < \infty$. *Then,*

(a) *if* $\delta_1 + \delta_2 > 0$,

$$\sqrt{N}\left(\widehat{\beta}_{BG} - \beta^0\right) \Longrightarrow N\left(0,\ \sigma_\mu^2 \Sigma_{\eta\eta}^{-1}\right).$$

(b) *if* $\delta_1 + \delta_2 = 0$,

$$\sqrt{N}\left(\widehat{\beta}_{BG} - \beta^0\right) \Longrightarrow N\left(0,\ \sigma_\mu^2\left[c^{-2\delta_1} D^0 M_P D^0 + \Sigma_{\eta\eta}\right]^{-1}\right).$$

Remark 4.3.

(i) Theorem 4.1 shows that, when $\delta_1 + \delta_2 \geq 1/2$, instrument weakness as modeled by the local-to-zero structure here is such that the 2*SLS* estimator becomes inconsistent. Moreover, the limiting distribution of 2*SLS* in these cases can be represented in terms of covariance matrix mixture of normals,

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

similar to the results obtained by Phillips (1989) and Choi and Phillips (1992) for the totally and partially unidentified cases and by Staiger and Stock (1997) within a local-to-zero framework for the purely cross-sectional and time series cases. Note further that only the randomness of the individual effects appears in the limiting distribution of 2*SLS*, while terms involving the idiosyncratic disturbances u_{it} and v_{it} are of a lower order. In consequence, even though we are considering a panel data setup where both the time and the cross-sectional dimensions are large, the behavior of the 2*SLS* actually bears a closer resemblance (than what one might have expected) to results in the purely cross-sectional cases, such as those which have been obtained by Phillips (1989), Choi and Phillips (1992), and Staiger and Stock (1997).

(ii) Note that the *BG* estimator comes from running OLS on the regression that is obtained from taking time averages of the structural equation (1), i.e., the regression

$$\overline{y}_{1,i} = \overline{Y}'_{2,i}\beta + \mu_i + \overline{u}_i,$$

where

$$\overline{y}_{1,i} = \frac{1}{T}\sum_{t=1}^{T} y_{1,it},\ \overline{Y}_{2,i} = \frac{1}{T}\sum_{t=1}^{T} Y_{2,it}, \text{ and } \overline{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}.$$

Moreover, from (2), we obtain

$$\overline{Y}_{2,i} = \Pi'\overline{Z}_i + \eta_i + \overline{v}_i,$$

with $\overline{Z}_i = T^{-1}\sum_{t=1}^{T} Z_{it}$ and $\overline{v}_i = T^{-1}\sum_{t=1}^{T} v_{it}$. Hence, an intuitive explanation for the consistency of the *BG* estimator is simply that by a law of large numbers for weakly dependent random sequences, $\overline{u}_i \approx 0$ and $\overline{v}_i \approx 0$ as $T \to \infty$, so that for large T

$$\overline{y}_{1,i} \approx \overline{Y}'_{2,i}\beta + \mu_i,$$

$$\overline{Y}_{2,i} \approx \Pi'\overline{Z}_i + \eta_i. \tag{10}$$

It follows that, asymptotically at least, $\overline{Y}_{2,i}$ is uncorrelated with the errors of the between-group regression (10) given the assumption that $\sigma_{\eta\mu} = 0$; and OLS estimation of this regression is, therefore consistent.

(iii) Although the *BG* estimator has more favorable asymptotic properties than the 2*SLS* estimator, it too has certain deficiencies in the present context. In particular, as can be seen from Theorem 4.2, its rate of convergence never exceeds $\sqrt{N}$, even when the underlying simultaneous equations model is strongly identified (i.e., $\delta_1 = \delta_2 = 0$). Hence, the precision with which it estimates seems to have been derived primarily from variations in the cross-sectional dimension, and it does not fully utilize the fact that the panel data model studied here has a large time dimension as well. Moreover, as we will

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

show in the next section of this paper, the BG estimator is inconsistent if the assumed covariance restriction $\sigma_{\eta\mu} = 0$ is invalid.

A goal of this paper is, thus, to construct an estimator, which remedies some of the problems of the 2SLS and the BG estimator. We start by considering a GMM estimator with criterion function given by

$$Q(\beta) = \widehat{g}(\beta)' \widehat{W} \widehat{g}(\beta), \tag{11}$$

where the vector of (sample) moment functions $\widehat{g}(\beta)$ can be written in partitioned form as

$$\underset{(G+K)\times 1}{\widehat{g}(\beta)} = \begin{bmatrix} \underset{G\times 1}{\widehat{g}_1(\beta)} \\ \underset{K\times 1}{\widehat{g}_2(\beta)} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} \overline{Y}_{2,i} \left(\overline{y}_{1,i} - \overline{Y}_{2,i}' \beta \right) \\ \sum_{i=1}^{N} \sum_{t=1}^{T} Z_{it} \left\{ \left(y_{1,it} - \overline{y}_{1,i} \right) - \left(Y_{2,it} - \overline{Y}_{2,i} \right)' \beta \right\} \end{bmatrix} \tag{12}$$

and where the weighting matrix $\widehat{W}$ is taken to be block diagonal and can be partitioned conformably with $\widehat{g}(\beta)$ as

$$\begin{aligned} \widehat{W} &= \begin{pmatrix} \widehat{W}_1 & 0 \\ 0 & \widehat{W}_2 \end{pmatrix} \\ &= \begin{pmatrix} \left[\left(1/\widehat{\sigma}_\mu^2 \right) \sum_{i=1}^{N} \overline{Y}_{2,i} \overline{Y}_{2,i}' \right]^{-1} & 0 \\ 0 & \left[\sum_{i=1}^{N} T \widehat{J}_{i,T} \right]^{-1} \end{pmatrix}, \end{aligned} \tag{13}$$

with $\widehat{\sigma}_\mu^2 = (1/N) \sum_{i=1}^{N} \left(\overline{y}_{1,i} - \overline{Y}_{2,i}' \overline{\beta} \right)^2$ for some preliminary estimator $\overline{\beta}$ to be defined below.

Note that the first set of moment functions $\widehat{g}_1(\beta)$ arises from the between-group regression. If we write

$$Q(\beta) = \widehat{g}_1(\beta)' \widehat{W}_1 \widehat{g}_1(\beta) + \widehat{g}_2(\beta)' \widehat{W}_2 \widehat{g}_2(\beta), \tag{14}$$

it is clear that the first part of the criterion function, i.e.,

$$\begin{aligned} \widehat{g}_1(\beta)' \widehat{W}_1 \widehat{g}_1(\beta) = &\left(\sum_{i=1}^{N} \overline{Y}_{2,i} \left(\overline{y}_{1,i} - \overline{Y}_{2,i}' \beta \right) \right)' \left[\left(1/\widehat{\sigma}_\mu^2 \right) \sum_{i=1}^{N} \overline{Y}_{2,i} \overline{Y}_{2,i}' \right]^{-1} \\ &\times \left(\sum_{i=1}^{N} \overline{Y}_{2,i} \left(\overline{y}_{1,i} - \overline{Y}_{2,i}' \beta \right) \right), \end{aligned}$$

is simply the objective function of the BG estimator, written in the present format to emphasize the fact that $\overline{Y}_{2,i}$ is (asymptotically) a valid instrument in this case. Alternatively, under Assumption 5, we can also consider using instruments based

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

on some estimate of the random effects η_i in the first-stage equation (2), say

$$\widehat{\eta}_i = \overline{Y}_{2,i} - \widehat{\Pi}'\overline{Z}_i,$$

where $\widehat{\Pi} = \left(Z'\left[I_N \otimes Q_{\iota_T}\right]Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right]Y_2$. However, it is not hard to see that asymptotically using $\widehat{\eta}_i$ as instruments gives the same information as $\overline{Y}_{2,i}$, so we will not consider this alternative here.

The second set of moment conditions $\widehat{g}_2(\beta)$ in (12) comes from the within-group regression. As we will see, the first part of the criterion function (14) plays a bigger role when instruments provided by coefficient restrictions are relatively weak and identification depends on the covariance restrictions, while the second part of (14) becomes more important in the case where the instruments Z_{it} are strong.

$\widehat{J}_{i,T}$ in equation (13) is the HAC estimator for the i^{th} cross sectional unit, so that

$$\widehat{J}_{i,T} = \sum_{\ell=0}^{T-1} w(\ell/q_T)\frac{1}{T}\sum_{t=\ell+1}^{T} \widehat{u}_{it}\widehat{u}_{it-\ell}\left(Z_{it} - \overline{Z}_i\right)\left(Z_{it-\ell} - \overline{Z}_i\right)' + \sum_{\ell=-T+1}^{-1} w(\ell/q_T)\frac{1}{T}\sum_{t=-\ell+1}^{T} \widehat{u}_{it}\widehat{u}_{it+\ell}\left(Z_{it+\ell} - \overline{Z}_i\right)\left(Z_{it} - \overline{Z}_i\right)', \qquad \textbf{(15)}$$

where $w(\cdot)$ is a real-valued kernel, q_T is a bandwidth parameter, and

$$\widehat{u}_{it} = \left(y_{1,it} - \overline{y}_{1,i}\right) - \left(Y_{2,it} - \overline{Y}_{2,i}\right)'\overline{\beta}.$$

To give conditions on the kernel function, we follow Parzen (1957) in defining the *characteristic exponent* of $w(x)$ to be the largest real number r such that

$$w_r = \lim_{x\to 0}\frac{1 - w(x)}{|x|^r}$$

exists and is finite. We make the following standard assumptions on the class of admissible kernels (cf., Hannan, 1970; Andrews,1991).

Assumption 7 (Kernel condition).

The kernel function $w(\cdot) : \mathbb{R} \to [-1, 1]$ is an even function which satisfies the following conditions

(a) $w(0) = 1$

(b) $\int_{-\infty}^{\infty} |w(x)|dx < \infty$;

(c) $w(\cdot)$ is continuous at zero and at all but a finite number of points,

(d) $w(\cdot)$ has characteristic exponent $r \geq 1$.

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Since the criterion function $Q(\beta)$ is quadratic in β, the GMM estimator $\widehat{\beta}_{GMM}$ obtained from minimizing (11) has an explicit representation given by

$$\begin{aligned}\widehat{\beta}_{GMM} &= \left[\frac{1}{\widehat{\sigma}_{\mu}^{2}}\sum_{i=1}^{N}\overline{Y}_{2,i}\overline{Y}_{2,i}'+\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{2,it}-\overline{Y}_{2,i}\right)Z_{it}'\widehat{W}_{2}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it}\left(Y_{2,it}-\overline{Y}_{2,i}\right)'\right]^{-1} \\ &\quad\times\left[\frac{1}{\widehat{\sigma}_{\mu}^{2}}\sum_{i=1}^{N}\overline{Y}_{2,i}\overline{y}_{1,i}+\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{2,it}-\overline{Y}_{2,i}\right)Z_{it}'\widehat{W}_{2}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{it}\left(y_{1,it}-\overline{y}_{1,i}\right)\right] \\ &= \left[\widehat{\sigma}_{\mu}^{-2}\frac{Y_{2}'\left[I_{N}\otimes P_{\iota_{T}}\right]Y_{2}}{T}+Y_{2}'\left[I_{N}\otimes Q_{\iota_{T}}\right]Z\widehat{W}_{2}Z'\left[I_{N}\otimes Q_{\iota_{T}}\right]Y_{2}\right]^{-1} \\ &\quad\times\left[\widehat{\sigma}_{\mu}^{-2}\frac{Y_{2}'\left[I_{N}\otimes P_{\iota_{T}}\right]y_{1}}{T}+Y_{2}'\left[I_{N}\otimes Q_{\iota_{T}}\right]Z\widehat{W}_{2}Z'\left[I_{N}\otimes Q_{\iota_{T}}\right]y_{1}\right], \qquad \textbf{(16)}\end{aligned}$$

where $\widehat{W}_{2}=\left[\sum_{i=1}^{N}T\widehat{J}_{i,T}\right]^{-1}$ as defined earlier. The HAC estimators $\widehat{J}_{i,T}$ $(i=1,\ldots,N)$ require a preliminary estimator for β. We can take this preliminary estimator $\overline{\beta}$ to be an "unweighted" version of (16), which does not use the correct HAC estimator as the weighting matrix, i.e.,

$$\begin{aligned}\overline{\beta} &= \left[\frac{Y_{2}'\left[I_{N}\otimes P_{\iota_{T}}\right]Y_{2}}{T}+Y_{2}'\left[I_{N}\otimes Q_{\iota_{T}}\right]Z\left(Z'\left[I_{N}\otimes Q_{\iota_{T}}\right]Z\right)^{-1}Z'\left[I_{N}\otimes Q_{\iota_{T}}\right]Y_{2}\right]^{-1} \\ &\quad\times\left[\frac{Y_{2}'\left[I_{N}\otimes P_{\iota_{T}}\right]y_{1}}{T}+Y_{2}'\left[I_{N}\otimes Q_{\iota_{T}}\right]Z\left(Z'\left[I_{N}\otimes Q_{\iota_{T}}\right]Z\right)^{-1}Z'\left[I_{N}\otimes Q_{\iota_{T}}\right]y_{1}\right].\end{aligned}$$

If covariance restrictions given by Assumption 5 hold, then it can be shown that $\overline{\beta}$ is consistent regardless of how weak the coefficient restrictions are. This result is stated in the following lemma.

LEMMA 4.4. *Suppose that Assumptions 1–6 hold and suppose that* $N,T\to\infty$ *such that* $N/T\to c$ *for some constant* c *with* $0<c<\infty$. *Then,*

$$\overline{\beta}\xrightarrow{p}\beta^{0},$$

for all $\delta_{1}\geq 0$ *and* $\delta_{2}\geq 0$.

We are now ready to state the main result of this section, which gives the rates of convergence and asymptotic distributions of the GMM estimator $\widehat{\beta}_{GMM}$ under varying degrees of instrument weakness. To facilitate the statement of our result, we introduce some additional notation. Let

$$\widehat{\Xi}_{1,NT}=\widehat{\sigma}_{\mu}^{-2}\frac{Y_{2}'\left[I_{N}\otimes P_{\iota_{T}}\right]Y_{2}}{T},$$

$$\widehat{\Xi}_{2,NT}=Y_{2}'\left[I_{N}\otimes Q_{\iota_{T}}\right]Z\widehat{W}_{2}Z'\left[I_{N}\otimes Q_{\iota_{T}}\right]Y_{2},$$

$$\widehat{\Xi}_{NT}=\widehat{\Xi}_{1,NT}+\widehat{\Xi}_{2,NT}.$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

THEOREM 4.5 (*GMM*). *Suppose that Assumptions 1–7 hold, and suppose that* $N, T \to \infty$ *such that* $N/T \to c$ *for some constant* c *with* $0 < c < \infty$ *and that* $q_T \to \infty$ *as* $T \to \infty$ *such that* $q_T^2/T \to 0$. *Then,*

(a) *if* $\delta_1 + \delta_2 > 1/2$,

$$\widehat{\Xi}_{NT}^{1/2}\left(\widehat{\beta}_{GMM} - \beta^0\right) = \widehat{\Xi}_{1,NT}^{1/2}\left(\widehat{\beta}_{GMM} - \beta^0\right) + o_p(1) \Longrightarrow N(0, I_G);$$

(b) *if* $\delta_1 + \delta_2 = 1/2$,

$$\widehat{\Xi}_{NT}^{1/2}\left(\widehat{\beta}_{GMM} - \beta^0\right) \Longrightarrow N(0, I_G);$$

(c) *if* $0 \leq \delta_1 + \delta_2 < 1/2$,

$$\widehat{\Xi}_{NT}^{1/2}\left(\widehat{\beta}_{GMM} - \beta^0\right) = \widehat{\Xi}_{2,NT}^{1/2}\left(\widehat{\beta}_{GMM} - \beta^0\right) + o_p(1) \Longrightarrow N(0, I_G).$$

Remark 4.6. From the proofs of Theorem 4.5 and the support lemmas (Lemmas S12–S14 given in the supplement to this paper), it is apparent that in the stronger instrument case where $0 \leq \delta_1 + \delta_2 < 1/2$, we have

$$T^{(1-\delta_1-\delta_2)}\left(\widehat{\beta}_{GMM} - \beta^0\right) = O_p(1),$$

so that the rate of convergence of $\widehat{\beta}_{GMM}$ in this case is $T^{\left(1-\delta_1-\delta_2\right)}$. This is a faster rate of convergence than that of the BG estimator $\widehat{\beta}_{BG}$, which is always $\sqrt{T}$ regardless of the strength of the instruments. The reason why the GMM estimator has a faster rate of convergence in the stronger instrument case is that it makes use of variations in both the cross-sectional and the time dimensions in this case, whereas the BG estimator always only makes use of variations along the cross-sectional dimension, as noted earlier in Remark 4.3(iii).

5. ESTIMATION WITHOUT COVARIANCE RESTRICTIONS

It is also of interest to analyze how the GMM estimator proposed in the last section would behave if the covariance restrictions stipulated in Assumption 5 do not hold. To study this case, we introduce the alternative assumption

Assumption 5*. $\sigma_{\eta\mu} = E\left[\eta_i \mu_i\right] \neq 0$ and $V\left(\eta_i \mu_i\right) = E\left[\mu_i^2 \eta_i \eta_i'\right] - \sigma_{\eta\mu}\sigma_{\eta\mu}' > 0$.

Note that in contrast to Assumption 5, Assumption 5* allows the random effect μ_i to be correlated with the explanatory variables $Y_{2,it}$ in the structural equation (1). Violation of Assumption 5 is possible, for example, for certain application of panel data methods to policy analysis and program evaluation where the policy variable may depend on certain time-invariant individual attribute that may also affect the outcome variable $y_{1,it}$. Results of the last section show that,

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

in situations where Assumption 5 is appropriate, our GMM estimator $\widehat{\beta}_{GMM}$ is consistent and asymptotically normal no matter how weak the available instruments may be. However, as we show below, these results no longer hold when there is correlation between the random effects μ_i and η_i, as specified in Assumption 5* above.

The next three results characterize the asymptotic properties of the between-group estimator $\widehat{\beta}_{BG}$, the preliminary estimator $\overline{\beta}$, and the GMM estimator $\widehat{\beta}_{GMM}$ when covariance restrictions are not assumed to hold.

THEOREM 5.1 (*BG*).

Suppose that Assumptions 1–4 and 5 hold and suppose that* $N, T \to \infty$ *such that* $N/T \to c$ *for some constant* c *with* $0 < c < \infty$. *Then,*

(a) *if* $\delta_1 + \delta_2 > 0$,

$$\widehat{\beta}_{BG} \xrightarrow{p} \beta^0 + \Sigma_{\eta\eta}^{-1}\sigma_{\eta\mu};$$

(b) *if* $\delta_1 = \delta_2 = 0$,

$$\widehat{\beta}_{BG} \xrightarrow{p} \beta^0 + \left[D^0 M_P D^0 + \Sigma_{\eta\eta}\right]^{-1}\sigma_{\eta\mu}.$$

LEMMA 5.2. *Suppose that Assumptions 1–4, 5*, and 6 hold and suppose that* $N, T \to \infty$ *such that* $N/T \to c$ *for some constant* c *with* $0 < c < \infty$. *Then,*

$$\overline{\beta} \xrightarrow{p} \beta^0,$$

if $0 \le \delta_1 + \delta_2 < 1/2$.

To help in characterizing the limiting behavior of the GMM estimator under Assumption 5*, we let

$$\overline{J}_{NT} = \frac{Z'\left(I_N \otimes Q_{\iota_T}\right)\Sigma_{u,NT}\left(I_N \otimes Q_{\iota_T}\right)Z}{NT},$$

and make the following additional assumption

Assumption 8. As $N, T \to \infty$ such that $N/T \to c$ for some constant c with $0 < c < \infty$,

$$\overline{J}_{NT} \to \overline{J}$$

for some $K \times K$ positive definite matrix $\overline{J}$.

THEOREM 5.3 (*GMM*).

Suppose that Assumptions 1–4, 5^*, *6–7, and 8 hold, and suppose that* $N, T \to \infty$ *such that* $N/T \to c$ *for some constant* c *with* $0 < c < \infty$ *and that* $q_T \to \infty$ *as* $T \to \infty$ *such that* $q_T^2/T \to 0$. *Then,*

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

(a) *if* $\delta_1 = \delta_2 = 0$,

$$\widehat{\Xi}_{NT}^{1/2}\left(\widehat{\beta}_{GMM} - \beta^0\right) \Longrightarrow N\left(\vartheta_1, I_G\right),$$

where

$$\vartheta_1 = c^{1/2+\delta_1}\left[D^{0\prime}M_Q\overline{J}^{-1}M_QD^0\right]^{-1/2}\left(\sigma_{\eta\mu}/\sigma_\mu^2\right),$$

(b) *if* $0 < \delta_1 + \delta_2 < 1/2$,

$$T^{1-2(\delta_1+\delta_2)}\left(\widehat{\beta}_{GMM} - \beta^0\right) \xrightarrow{p} \vartheta_2,$$

where

$$\vartheta_2 = c^{2\delta_1}\left[D^{0\prime}M_Q\overline{J}^{-1}M_QD^0\right]^{-1}\left(\sigma_{\eta\mu}/\sigma_\mu^2\right).$$

Remark 5.4.

(i) Note that Assumption 8 requires that the double-indexed matrix sequence $\overline{J}_{NT}$ converges to a limit. The basic qualitative features of the limiting behavior of $\widehat{\beta}_{GMM}$ (i.e., whether it is consistent or inconsistent, whether it has a second order bias, etc.) are not changed by this added requirement. This extra condition, however, does help us to write down cleaner formulae for characterizing the asymptotic properties of $\widehat{\beta}_{GMM}$.

(ii) Lemma 5.2 shows that as long as the instruments are not too weak, the preliminary estimator $\overline{\beta}$ will be consistent under Assumption 5* as well. This allows us to obtain a consistent estimator of the long-run covariance matrix, provided that the condition $0 \leq \delta_1 + \delta_2 < 1/2$ is satisfied.

(iii) In the weaker instrument case where $\delta_1 + \delta_2 \geq 1/2$, $\widehat{\beta}_{GMM}$ is no longer consistent, but its asymptotic behavior is harder to characterize, since the HAC estimator in this case is also inconsistent.

(iv) Even in the stronger instrument case where $0 \leq \delta_1 + \delta_2 < 1/2$, $\widehat{\beta}_{GMM}$ suffers from second order biases, as can be seen from the results reported in Theorem 5.3, so that it is difficult to conduct statistical inference with this estimator when the covariance restrictions do not hold. However, it turns out that it is possible to define a new estimator, which will correct for this bias when the covariance restrictions are violated but which, on the other hand, will also be asymptotically equivalent to the GMM estimator when bias correction is unnecessary. We construct such an estimator below.

To proceed, let χ denote a chi-square random variable with G degrees of freedom, i.e., $\chi \sim \chi_G^2$. Also, define

$$\mathbb{T}_N = \widetilde{\sigma}_{\eta\mu}'\widehat{V}^{-1}\widetilde{\sigma}_{\eta\mu}, \tag{17}$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

where $\widetilde{\sigma}_{\eta\mu} = N^{-1}\sum_{i=1}^{N}\widetilde{\eta}_i\widetilde{\mu}_i$ and $\widetilde{V} = N^{-1}\{N^{-1}\sum_{i=1}^{N}\widetilde{\mu}_i^2\widetilde{\eta}_i\widetilde{\eta}_i' - \widetilde{\sigma}_{\eta\mu}\widetilde{\sigma}_{\eta\mu}'\}$ with $\widetilde{\eta}_i = \overline{Y}_{2,i} - \widehat{\Pi}'\overline{Z}_i$, $\widetilde{\mu}_i = \overline{y}_{1,i} - \overline{Y}_{2,i}'\widehat{\beta}_{GMM}$, and $\widehat{\Pi} = \left(Z'\left[I_N \otimes Q_{\iota_T}\right]Z\right)^{-1}Z'\left[I_N \otimes Q_{\iota_T}\right]Y_2$; so that $\mathbb{T}_N$ is a test statistic for testing the null hypothesis $H_0 : \sigma_{\eta\mu} = 0$. Now, consider the estimator

$$\widetilde{\beta}_C = p_\chi\widetilde{\beta}_{GMM} + \left(1 - p_\chi\right)\left(\widetilde{\beta}_{GMM} - BC\right), \tag{18}$$

where $p_\chi = P\left(\chi > \mathbb{T}_N - G\ln N\right)$ is the (approximate) p-value of the model selection criterion $\mathbb{T}_N - G\ln N$ and where the bias-correction factor BC takes the form

$$BC = \left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right]Y_2}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N \otimes Q_{\iota_T}\right]Y_2\right]^{-1}\left[N\widetilde{\sigma}_{\eta\mu}/\widetilde{\sigma}_\mu^2\right],$$

with $\widetilde{\sigma}_\mu^2 = N^{-1}\sum_{i=1}^{N}\widetilde{\mu}_i^2$. In addition,

$$\widetilde{\beta}_{GMM} = \left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right]Y_2}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N \otimes Q_{\iota_T}\right]Y_2\right]^{-1} \times \left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right]y_1}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N \otimes Q_{\iota_T}\right]y_1\right],$$

so that it is similar to the GMM estimator $\widehat{\beta}_{GMM}$ defined earlier, except that it makes use of the alternative weighting matrix $\widetilde{W}_2 = \left[\sum_{i=1}^{N}T\widetilde{J}_{i,T}\right]^{-1}$ with HAC estimator $\widetilde{J}_{i,T}$ similar to (15) but based on residuals formed from $\widehat{\beta}_{GMM}$, i.e.,

$$\widetilde{u}_{it} = \left(y_{1,it} - \overline{y}_{1,i}\right) - \left(Y_{2,it} - \overline{Y}_{2,i}\right)'\widehat{\beta}_{GMM},$$

instead of residuals formed from the preliminary estimator $\overline{\beta}$.

Note that the combined (or model averaging) estimator defined by (18) is a smoothed version of a pretest estimator, where smoothing of the weight function here is accomplished by a p-value transformation. Since the magnitude of the p-value provides important information on how strongly the data contradicts the null hypothesis, it seems that using the p-value transformation as a mechanism for smoothing is quite natural if we were to approach estimator combination from a purely frequentist perspective. On the other hand, an alternative way of smoothing the weight function which has a more Bayesian motivation can be obtained by using a logistic transformation of $\widetilde{\Delta}_{BIC,N} = \mathbb{T}_N - G\ln N$, which is approximately the difference of the values of the Schwarz criterion (or BIC) under the null and alternative hypotheses. Such a weight function results in a model averaging estimator of the form

$$\widetilde{\beta}_{SBIC} = w_{BIC}\widetilde{\beta}_{GMM} + (1 - w_{BIC})\left(\widetilde{\beta}_{GMM} - BC\right), \tag{19}$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

where

$$w_{BIC} = \frac{\exp\left\{-\frac{1}{2}\widetilde{\Delta}_{BIC,N}\right\}}{1+\exp\left\{-\frac{1}{2}\widetilde{\Delta}_{BIC,N}\right\}}.$$

A weighted average estimator of the type given in (19) has been introduced in other contexts by Buckland, Burnham, and Augustin (1997) and is known in the literature on model selection and model averaging as the smoothed BIC estimator.[1] Other important papers that have studied estimators of this type as well as other model averaging estimators include Hjort and Claeskens (2003) and Hansen (2007). Here, we add to this literature by studying the asymptotic properties of both $\widetilde{\beta}_C$ and $\widetilde{\beta}_{SBIC}$ when there is uncertainty about the strength and validity of alternative identifying restrictions. To proceed, define

$$\widetilde{\Xi}_{1,NT} = \widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right]Y_2}{T},$$

$$\widetilde{\Xi}_{2,NT} = Y_2'\left[I_N \otimes Q_{\iota_T}\right]Z\widetilde{W}_2 Z'\left[I_N \otimes Q_{\iota_T}\right]Y_2,$$

$$\widetilde{\Xi}_{NT} = \widetilde{\Xi}_{1,NT} + \widetilde{\Xi}_{2,NT}.$$

The next two theorems give our main results on the model averaging estimators $\widetilde{\beta}_C$ and $\widetilde{\beta}_{SBIC}$.

THEOREM 5.5. *Suppose that Assumptions 1–7 hold, and suppose that* $N, T \to \infty$ *such that* $N/T \to c$ *for some constant* c *with* $0 < c < \infty$ *and that* $q_T \to \infty$ *as* $T \to \infty$ *such that* $q_T^2/T \to 0$. *Then,*

(a) *if* $\delta_1 + \delta_2 > 1/2$,

$$\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_C - \beta^0\right) = \widetilde{\Xi}_{1,NT}^{1/2}\left(\widetilde{\beta}_C - \beta^0\right) + o_p(1) \Longrightarrow N(0, I_G)$$

and

$$\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_{SBIC} - \beta^0\right) = \widetilde{\Xi}_{1,NT}^{1/2}\left(\widetilde{\beta}_{SBIC} - \beta^0\right) + o_p(1) \Longrightarrow N(0, I_G);$$

(b) *if* $\delta_1 + \delta_2 = 1/2$,

$$\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_C - \beta^0\right) \Longrightarrow N(0, I_G),$$

and

$$\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_{SBIC} - \beta^0\right) \Longrightarrow N(0, I_G);$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

(c) *if* $0 \leq \delta_1 + \delta_2 < 1/2$,

$$\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_C - \beta^0\right) = \widetilde{\Xi}_{2,NT}^{1/2}\left(\widetilde{\beta}_C - \beta^0\right) + o_p(1) \Longrightarrow N(0, I_G)$$

and

$$\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_{SBIC} - \beta^0\right) = \widetilde{\Xi}_{2,NT}^{1/2}\left(\widetilde{\beta}_{SBIC} - \beta^0\right) + o_p(1) \Longrightarrow N(0, I_G).$$

THEOREM 5.6. *Suppose that Assumptions 1–4, 5*, 6–7 hold, and suppose that* $N, T \to \infty$ *such that* $N/T \to c$ *for some constant* c *with* $0 < c < \infty$ *and that* $q_T \to \infty$ *as* $T \to \infty$ *such that* $q_T^2/T \to 0$. *Furthermore, suppose that* $0 \leq \delta_1 + \delta_2 < 1/2$*; then,*

$$\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_C - \beta^0\right) = \widetilde{\Xi}_{2,NT}^{1/2}\left(\widetilde{\beta}_C - \beta^0\right) + o_p(1) \Longrightarrow N(0, I_G)$$

and

$$\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_{SBIC} - \beta^0\right) = \widetilde{\Xi}_{2,NT}^{1/2}\left(\widetilde{\beta}_{SBIC} - \beta^0\right) + o_p(1) \Longrightarrow N(0, I_G).$$

Remark 5.7.

(i) Note that $\widetilde{\beta}_C$ and $\widetilde{\beta}_{SBIC}$ are asymptotically equivalent in the cases studied in Theorems 5.5 and 5.6. Indeed, both $\widetilde{\beta}_C$ and $\widetilde{\beta}_{SBIC}$ have a certain robustness quality. In particular, they are consistent and asymptotically normal when the covariance restrictions hold no matter how weak are the instruments provided by coefficient restrictions. On the other hand, they are also consistent and asymptotically normal when the covariance restrictions do not hold, provided that one has relatively strong instruments. Moreover, for both estimators, the same standardization factor $\widetilde{\Xi}_{NT}^{1/2}$ applies across these various cases, so that one can conduct inference without explicit knowledge of which parameter sequence happens to be the most appropriate. Of course, consistency breaks down when neither the covariance restrictions nor the coefficient restrictions are valid, but this is the case where one does not expect to have reliable point estimation.

(ii) Note that p_χ is defined to be the "p-value" associated with the model selection criterion $\mathbb{T}_N - G \ln N$ and not the (asymptotic) p-value of the test statistic $\mathbb{T}_N$. This is important for obtaining the results above, since in this case one can show that $p_\chi \xrightarrow{p} 1$ under $H_0 : \sigma_{\eta\mu} = 0$; so that, when the null is true, $\widetilde{\beta}_C$ is asymptotically equivalent to $\widehat{\beta}_{GMM}$. On the other hand, under the alternative hypothesis $H_1 : \sigma_{\eta\mu} \neq 0$; we have that $p_\chi \xrightarrow{p} 0$, so that asymptotically under the alternative, full weight is given to the bias-corrected estimator $\widetilde{\beta}_{GMM} - BC$.

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

6. MONTE CARLO STUDY

This section reports the results of a Monte Carlo study, which compares a number of estimators on the basis of bias and mean-squared error (MSE). The data generating processes used in the Monte Carlo design are stripped-down versions of the model given in equations (1)–(2), i.e.,

$$y_{1,it} = \beta^0 y_{2,it} + \mu_i + u_{it}, \tag{20}$$

$$y_{2,it} = z_{it}\pi + \eta_i + v_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T; \tag{21}$$

so that the structural equation (20) here only has one endogenous regressor ($G = 1$). To keep the design simple, we also consider only the case with a single instrument $z_{it} \equiv i.i.d.N(0.5, 1)$; and, hence, $K = 1$ for all results reported below. Moreover, we take $\beta^0 = 1$, $N = 100$, and $T = 100$.

The individual effects and the idiosyncratic errors are assumed to be normally distributed. More specifically, we take

$$\begin{pmatrix} \mu_i \\ \eta_i \end{pmatrix} \equiv i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix} \right),$$

$$\begin{pmatrix} u_{it} \\ v_{it} \end{pmatrix} \equiv i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

We vary $\theta = (0, 0.25, 0.5)$ and $\rho = (0, 0.25, 0.5)$. To control for instrument weakness, we also choose π, the first-stage coefficient on the instrument, so that the population R^2 on the first-stage equation equals 0.1, 0.01, and 0.001.

In addition to $\widetilde{\beta}_{GMM}$, $\widehat{\beta}_{2SLS}$, $\widetilde{\beta}_{C}$, and $\widetilde{\beta}_{SBIC}$ discussed in earlier sections of this paper, we also consider the following smoothed AIC estimator in our comparison of estimators:

$$\widetilde{\beta}_{SAIC} = \frac{\exp\left\{-(1/2)\,\widetilde{\Delta}_{AIC}\right\}}{1+\exp\left\{-(1/2)\,\widetilde{\Delta}_{AIC}\right\}}\widetilde{\beta}_{GMM} + \frac{1}{1+\exp\left\{-(1/2)\,\widetilde{\Delta}_{AIC}\right\}}\left(\widetilde{\beta}_{GMM} - BC\right),$$

where $\mathbb{T}_N$ is the test statistic defined in (17) above and $\widetilde{\Delta}_{AIC} = \mathbb{T}_N - 2$. Note that, of course, $\widetilde{\beta}_{SAIC}$ has a form very similar to $\widetilde{\beta}_{SBIC}$, except that the weight functions are constructed using AIC as opposed to BIC.

Tables 1–6 present the results of our simulation study. The odd number tables report results on bias whereas the even number tables give the results for MSE. All numbers reported are averages computed from 10,000 simulation runs.

Overall, the GMM estimator $\widetilde{\beta}_{GMM}$ and the model averaging estimators $\widetilde{\beta}_{C}$, $\widetilde{\beta}_{SAIC}$, and $\widetilde{\beta}_{SBIC}$ outperform the $2SLS$ estimator $\widehat{\beta}_{2SLS}$ in terms of both bias and MSE. The efficiency gain from using $\widetilde{\beta}_{GMM}$ or one of the model averaging

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

TABLE 1. Estimator Bias, $\rho = 0$

R^2	θ	$\widetilde{\beta}_{GMM}$	$\widehat{\beta}_{2SLS}$	$\widetilde{\beta}_{C}$	$\widetilde{\beta}_{SAIC}$	$\widetilde{\beta}_{SBIC}$
	0	−0.0003959	0.0004190	−0.0003940	−0.0003921	−0.0003895
0.1	0.25	0.02013	−0.004214	0.008506	0.004260	0.007357
	0.50	0.04254	−0.009067	0.004048	0.003960	0.004292
	0	0.0004867	−0.03690	0.0004468	0.0007944	0.0006127
0.01	0.25	0.1266	−0.1711	0.1184	0.09077	0.1066
	0.50	0.2822	−0.07286	0.2106	0.1811	0.2023
	0	−0.001653	2.033	−0.001653	−0.001629	−0.001649
0.001	0.25	0.2254	−0.5470	0.2253	0.2194	0.2232
	0.50	0.4631	1.318	0.4630	0.4534	0.4595

***Results are based on 10,000 simulations.

TABLE 2. Estimator MSE, $\rho = 0$

R^2	θ	$\widetilde{\beta}_{GMM}$	$\widehat{\beta}_{2SLS}$	$\widetilde{\beta}_{C}$	$\widetilde{\beta}_{SAIC}$	$\widetilde{\beta}_{SBIC}$
	0	0.0008403	0.01640	0.0008444	0.0008598	0.0008467
0.1	0.25	0.001245	0.01668	0.001055	0.0009400	0.0009913
	0.50	0.002709	0.01722	0.0009338	0.0009285	0.0009403
	0	0.004944	5.643	0.004965	0.005211	0.005013
0.01	0.25	0.02155	39.95	0.02049	0.01494	0.01770
	0.50	0.08599	74.39	0.05727	0.04237	0.05148
	0	0.009294	10770	0.009294	0.009358	0.009301
0.001	0.25	0.05989	8132	0.05989	0.05745	0.05899
	0.50	0.2225	3319	0.2225	0.2140	0.2194

***Results are based on 10,000 simulations.

TABLE 3. Estimator Bias, $\rho = 0.25$

R^2	θ	$\widetilde{\beta}_{GMM}$	$\widehat{\beta}_{2SLS}$	$\widetilde{\beta}_{C}$	$\widetilde{\beta}_{SAIC}$	$\widetilde{\beta}_{SBIC}$
	0	0.00002349	0.0003210	−0.00002196	−0.00009489	−0.00005134
0.1	0.25	0.02028	−0.005050	0.008493	0.004271	0.007336
	0.50	0.04173	−0.009257	0.003674	0.003595	0.003901
	0	0.002468	0.01090	0.002533	0.002165	0.002395
0.01	0.25	0.1247	−0.1046	0.1161	0.08826	0.1042
	0.50	0.2675	−0.5217	0.1843	0.1604	0.1795
	0	0.002958	−0.3556	0.002958	0.003480	0.003156
0.001	0.25	0.2282	0.3781	0.2282	0.2217	0.2259
	0.50	0.4612	−0.0163	0.4612	0.4500	0.4571

***Results are based on 10,000 simulations.

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

TABLE 4. Estimator MSE, $\rho = 0.25$

R^2	θ	$\widetilde{\beta}_{GMM}$	$\widehat{\beta}_{2SLS}$	$\widetilde{\beta}_{C}$	$\widetilde{\beta}_{SAIC}$	$\widetilde{\beta}_{SBIC}$
	0	0.00008424	0.01636	0.0008481	0.0008650	0.0008512
0.1	0.25	0.001219	0.01675	0.001033	0.0009232	0.0009714
	0.50	0.002572	0.01724	0.0009182	0.0009136	0.0009235
	0	0.005046	3.871	0.005081	0.005369	0.005145
0.01	0.25	0.02032	26.57	0.01906	0.01357	0.01629
	0.50	0.07667	1429	0.04422	0.03328	0.04058
	0	0.009261	2895	0.009261	0.009355	0.009278
0.001	0.25	0.06061	1769	0.06061	0.05784	0.05959
	0.50	0.2203	3377	0.2203	0.2105	0.2166

***Results are based on 10,000 simulations.

TABLE 5. Estimator Bias, $\rho = 0.5$

R^2	θ	$\widetilde{\beta}_{GMM}$	$\widehat{\beta}_{2SLS}$	$\widetilde{\beta}_{C}$	$\widetilde{\beta}_{SAIC}$	$\widetilde{\beta}_{SBIC}$
	0	0.0001211	0.001331	0.00003560	−0.0001851	−0.00004909
0.1	0.25	0.02016	−0.003385	0.008279	0.004124	0.007156
	0.50	0.04095	−0.009454	0.003310	0.003241	0.003522
	0	0.003146	−0.1651	0.003222	0.002395	0.002904
0.01	0.25	0.1224	−0.08286	0.1118	0.08357	0.09987
	0.50	0.2493	0.7102	0.1542	0.1369	0.1529
	0	0.007806	0.7249	0.007806	0.008784	0.008181
0.001	0.25	0.2298	1.143	0.2298	0.2233	0.2275
	0.50	0.4561	0.1443	0.4560	0.4426	0.4511

***Results are based on 10,000 simulations.

TABLE 6. Estimator MSE, $\rho = 0.5$

R^2	θ	$\widetilde{\beta}_{GMM}$	$\widehat{\beta}_{2SLS}$	$\widetilde{\beta}_{C}$	$\widetilde{\beta}_{SAIC}$	$\widetilde{\beta}_{SBIC}$
	0	0.0008353	0.01604	0.0008425	0.0008580	0.0008450
0.1	0.25	0.001189	0.0009837	0.001026	0.0009207	0.0009634
	0.50	0.002444	0.01769	0.0009043	0.0009006	0.0009087
	0	0.005054	81.88	0.005064	0.005249	0.005082
0.01	0.25	0.01915	26.20	0.01743	0.01220	0.01478
	0.50	0.06598	7284	0.03146	0.02460	0.02986
	0	0.009226	8380	0.009226	0.009325	0.009246
0.001	0.25	0.06092	4123	0.06092	0.05801	0.05986
	0.50	0.2150	1422	0.2150	0.2032	0.2106

***Results are based on 10,000 simulations.

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

estimators relative to $\widehat{\beta}_{2SLS}$ is especially apparent in the tables, as the MSE of $\widehat{\beta}_{2SLS}$ tends to be considerably larger than that of any of the other estimators considered here. This is in accord with our theory since, like the BG estimator, the precision with which $2SLS$ estimates is derived primarily from variations in the cross-sectional dimension whereas GMM and the model averaging estimators all make better use of additional variations in the time dimension.

None of the model averaging estimators $\widetilde{\beta}_C$, $\widetilde{\beta}_{SAIC}$, and $\widetilde{\beta}_{SBIC}$ uniformly dominate the others in performance. In particular, $\widetilde{\beta}_C$ and $\widetilde{\beta}_{SBIC}$ tend to have a bit lower bias and MSE than $\widetilde{\beta}_{SAIC}$ when $\theta = \sigma_{\eta\mu} = 0$; but the latter is the slightly better estimator when $\theta \neq 0$. This is likely due to the fact that the weight function of $\widetilde{\beta}_C$ and $\widetilde{\beta}_{SBIC}$ depends on BIC, which penalizes the higher-dimensional model (here, the alternative hypothesis $\sigma_{\eta\mu} \neq 0$) more strongly than AIC. Hence, in finite sample, $\widetilde{\beta}_C$ and $\widetilde{\beta}_{SBIC}$ would put greater weight on $\widetilde{\beta}_{GMM}$ (as opposed to the bias-corrected GMM estimator) than $\widetilde{\beta}_{SAIC}$ and, in consequence, perform better in scenarios where $\widetilde{\beta}_{GMM}$ may be more appropriate.

Finally, note that the model averaging estimators are slightly less bias than the GMM estimator $\widetilde{\beta}_{GMM}$, particularly when $\theta \neq 0$ and $R^2 = 0.1$ or $R^2 = 0.01$. This is consistent with our theory, since in this case we would expect $\widetilde{\beta}_{GMM}$ to have a second-order bias whereas the model averaging estimators all have a component which tries to correct for this bias. Moreover, the advantage of the model averaging estimators is less noticeable when $R^2 = 0.001$, because the bias correction mechanism becomes less effective as the quality of the instruments deteriorates.

7. CONCLUDING REMARKS

This paper considers the estimation of a panel data simultaneous equations model under both coefficient and covariance restrictions but in a situation where one or the other set of identifying conditions may not be valid or may hold only weakly. The main innovation of the paper is the construction of two new estimators, both of which are weighted-averages of a GMM estimator and a bias-corrected GMM estimator. The main difference between the two proposed estimators lies with their weighting function, with one being based on an approximate p-value transformation of BIC while the other uses a logistic transformation. We show that these new estimators have certain robustness properties against possible failure of either the coefficient restrictions or the covariance restrictions. We also report the results of a Monte Carlo study, which documents the efficiency gains of the GMM estimator and of the model averaging estimators over $2SLS$. The basic idea of this paper can be extended to more general setting where there might be uncertainty with regard to alternative sets of identifying restrictions and where estimation robustness with respect to possible failure of some of the identifying restrictions is desired. We intend to pursue extensions of this work in future studies.

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

NOTE

1. We thank the Guest Editor, Bruce Hansen, for suggesting the smoothed BIC estimator and also the smoothed AIC estimator, whose finite sample properties we study via Monte Carlo simulation in Section 6 of the paper.

REFERENCES

Alonso-Borrego, C. & M. Arellano (1999) Symmetrically normalized instrumental-variable estimation using panel data. *Journal of Business and Economic Statistics* 17, 36–49.

Anderson, T.W. (1971). *The Statistical Analysis of Time Series.* John Wiley & Sons.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Baltagi, B.H. (1981) Simultaneous equations with error components. *Journal of Econometrics* 17, 189–200.

Baltagi, B.H. & Q. Li (1992) A note on the estimation of simultaneous equations with error components. *Econometric Theory* 8, 113–119.

Buckland, S.T., K.P. Burnham, & N.H. Augustin (1997) Model selection: An integral part of inference. *Biometrics* 53, 603–618.

Bun, M. & F. Windmeijer (2007) The weak instrument problem of the system GMM estimator in dynamic panel data models. CEMMP working paper.

Chao, J.C. & N.R. Swanson (2005) Consistent estimation with a large number of weak instruments. *Econometrica* 73, 1673–1692.

Choi, I. & P.C.B. Phillips (1992) Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. *Journal of Econometrics* 51, 113–150.

Han, C. & P.C.B. Phillips (2006) GMM with many moment conditions. *Econometrica* 74, 147–192.

Hannan, E.J. (1970) *Multiple Time Series*. John Wiley & Sons.

Hansen, B.E. (2007) Least squares model averaging. *Econometrica* 75, 1175–1189.

Hausman, J.A. & W.E. Taylor (1983) Identification in linear simultaneous equations models with covariance restrictions: An instrumental variables interpretation. *Econometrica* 51, 1527–1549.

Hausman, J.A., W.K. Newey, & W.E. Taylor (1987) Efficient estimation and identification of simultaneous equation models with covariance restrictions. *Econometrica* 55, 849–874.

Hjort, N.L. & G. Claeskens (2003) Frequentist model average estimators. *Journal of the American Statistical Association* 98, 879–899.

Hsiao, C. (2003) *Analysis of Panel Data*, 2nd ed. Cambridge University Press.

Liu, T.-C. (1960) Underidentification, structural estimation, and forecasting. *Econometrica* 28, 855–865.

Parzen, E. (1957) On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28, 329–348.

Phillips, P.C.B. (1980) The exact distribution of instrumental variable estimators in an equation containing n+1 endogenous variables. *Econometrica* 48, 861–878.

Phillips, P.C.B. (1984) The exact distribution of LIML: I. *International Economic Review* 25, 249–261.

Phillips, P.C.B. (1985) The exact distribution of LIML: II. *International Economic Review* 26, 21–36.

Phillips, P.C.B. (1989) Partially identified econometric models. *Econometric Theory* 5, 181–240.

Phillips, P.C.B. (2006) A remark on bimodality and weak instrumentation in structural equation estimation. *Econometric Theory* 22, 947–960.

Prucha, I.R. (1985) Maximum likelihood and instrumental variable estimation in simultaneous equations systems with error components. *International Economic Review* 26, 491–506.

Sims, C.A. (1980) Macroeconomics and reality. *Econometrica* 48, 1–48.

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Staiger, D. & J.H. Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.

Stock, J.H. & J.H. Wright (2000) GMM with weak identification. *Econometrica* 68, 1055–1096.

Stock, J.H. & M. Yogo (2005) Asymptotic distributions of instrumental variables statistics and many weak instruments. In D.W.K.Andrews & J.H.Stock (eds.), *Festschrift in Honor of Thomas Rothenberg*. Cambridge University Press.

APPENDIX

The proofs given in this appendix rely on a number of lemmas, which are stated and proved in a supplement to this paper. Lemmas given in the supplement are numbered as Lemma S1, Lemma S2, etc.; which allow them to be easily identified when referenced below. Also, throughout this appendix, we let C denote a generic positive constant that may be different in different uses.

Proof of Theorem 4.1.

To proceed, first write

$$\widehat{\beta}_{2SLS} - \beta^0 = (Y_2' P_Z Y_2)^{-1} Y_2' P_Z \left[(I_N \otimes \iota_T) \mu + u \right]. \tag{22}$$

Focusing first on the "denominator" of (22), we have that

$$\begin{aligned} Y_2' P_Z Y_2 = {} & \frac{D^{0\prime} Z' Z D^0}{N^{2\delta_1} T^{2\delta_2}} + \left(\sum_{i=1}^N \eta_i \overline{Z}_i' \right) \left(\frac{Z'Z}{T^2} \right)^{-1} \left(\sum_{i=1}^N \overline{Z}_i \eta_i' \right) \\ & + \left(\sum_{i=1}^N \sum_{t=1}^T v_{it} Z_{it}' \right) (Z'Z)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T Z_{it} v_{it}' \right) \frac{T^{1-\delta_2}}{N^{\delta_1}} D^{0\prime} \left(\sum_{i=1}^N \overline{Z}_i \eta_i' \right) \\ & + \left(\sum_{i=1}^N \eta_i \overline{Z}_i' \right) D^0 \frac{T^{1-\delta_2}}{N^{\delta_1}} + \frac{1}{N^{\delta_1} T^{\delta_2}} D^{0\prime} \left(\sum_{i=1}^N \sum_{t=1}^T Z_{it} v_{it}' \right) \\ & + \left(\sum_{i=1}^N \sum_{t=1}^T v_{it} Z_{it}' \right) D^0 \frac{1}{N^{\delta_1} T^{\delta_2}} + \left(\sum_{i=1}^N \sum_{t=1}^T v_{it} Z_{it}' \right) \left(\frac{Z'Z}{T} \right)^{-1} \left(\sum_{i=1}^N \overline{Z}_i \eta_i' \right) \\ & + \left(\sum_{i=1}^N \eta_i \overline{Z}_i' \right) \left(\frac{Z'Z}{T} \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T Z_{it} v_{it}' \right). \end{aligned}$$

Similarly, for the "numerator" of (22), we have

$$\begin{aligned} & Y_2' P_Z \left[(I_N \otimes \iota_T) \mu + u \right] \\ & \quad = \frac{T^{1-\delta_2}}{N^{\delta_1}} D^{0\prime} \left(\sum_{i=1}^N \overline{Z}_i \mu_i \right) + \frac{1}{N^{\delta_1} T^{\delta_2}} D^{0\prime} Z' u + \left(\sum_{i=1}^N \eta_i \overline{Z}_i' \right) \left(\frac{Z'Z}{T} \right)^{-1} Z' u \\ & \qquad + \left(\sum_{i=1}^N \sum_{t=1}^T v_{it} Z_{it}' \right) \left(\frac{Z'Z}{T} \right)^{-1} \left(\sum_{i=1}^N \overline{Z}_i \mu_i \right) + \left(\sum_{i=1}^N \sum_{t=1}^T v_{it} Z_{it}' \right) (Z'Z)^{-1} Z' u \\ & \qquad + \left(\sum_{i=1}^N \eta_i \overline{Z}_i' \right) \left(\frac{Z'Z}{T^2} \right)^{-1} \left(\sum_{i=1}^N \overline{Z}_i \mu_i \right). \end{aligned}$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Now, applying Assumptions 1 and 2, parts (c) and (d) of Lemma S1, part (a) of Lemma S4, and Lemma S7 (given in the supplement), we obtain

$$\begin{aligned}\frac{D^{0\prime}Z'ZD^0}{N^{2\delta_1}T^{2\delta_2}} &= N^{1-2\delta_1}T^{1-2\delta_2}\left[\frac{D^{0\prime}Z'\left[I_N\otimes P_{\iota_T}\right]ZD^0}{NT}+\frac{D^{0\prime}Z'\left[I_N\otimes Q_{\iota_T}\right]ZD^0}{NT}\right]\\ &= O_p\left(T^{2(1-\delta_1-\delta_2)}\right),\end{aligned}$$

$$\begin{aligned}\left(\sum_{i=1}^N \eta_i\overline{Z}_i'\right)\left(\frac{Z'Z}{T^2}\right)^{-1}\left(\sum_{i=1}^N \overline{Z}_i\eta_i'\right) &= O_p\left(\sqrt{N}\right)O_p\left(1\right)O_p\left(\sqrt{N}\right)\\ &= O_p\left(N\right)=O_p\left(T\right),\\ \left(\sum_{i=1}^N\sum_{t=1}^T v_{it}Z_{it}'\right)\left(Z'Z\right)^{-1}\left(\sum_{i=1}^N\sum_{t=1}^T Z_{it}v_{it}'\right) &= O_p\left(\sqrt{NT}\right)O_p\left(N^{-1}T^{-1}\right)O_p\left(\sqrt{NT}\right)\\ &= O_p\left(1\right),\\ \frac{T^{1-\delta_2}}{N^{\delta_1}}D^{0\prime}\left(\sum_{i=1}^N \overline{Z}_i\eta_i'\right) &= O_p\left(N^{-\delta_1}T^{1-\delta_2}\right)O_p\left(\sqrt{N}\right)\\ &= O_p\left(T^{(3/2-\delta_1-\delta_2)}\right)\\ \frac{1}{N^{\delta_1}T^{\delta_2}}D^{0\prime}\left(\sum_{i=1}^N\sum_{t=1}^T Z_{it}v_{it}'\right) &= O_p\left(N^{-\delta_1}T^{-\delta_2}\right)O_p\left(\sqrt{NT}\right)\\ &= O_p\left(T^{(1-\delta_1-\delta_2)}\right),\\ \left(\sum_{i=1}^N\sum_{t=1}^T v_{it}Z_{it}'\right)\left(\frac{Z'Z}{T}\right)^{-1}\left(\sum_{i=1}^N \overline{Z}_i\eta_i'\right) &= O_p\left(\sqrt{NT}\right)O_p\left(T^{-1}\right)O_p\left(\sqrt{N}\right)\\ &= O_p\left(\sqrt{T}\right),\end{aligned}$$

and

$$\begin{aligned}\frac{T^{1-\delta_2}}{N^{\delta_1}}D^{0\prime}\left(\sum_{i=1}^N \overline{Z}_i\mu_i\right) &= O_p\left(N^{-\delta_1}T^{1-\delta_2}\right)O_p\left(\sqrt{N}\right)\\ &= O_p\left(T^{(3/2-\delta_1-\delta_2)}\right),\\ \frac{1}{N^{\delta_1}T^{\delta_2}}D^{0\prime}Z'u &= O_p\left(N^{-\delta_1}T^{-\delta_2}\right)O_p\left(\sqrt{NT}\right)\\ &= O_p\left(T^{(1-\delta_1-\delta_2)}\right),\\ \left(\sum_{i=1}^N \eta_i\overline{Z}_i'\right)\left(\frac{Z'Z}{T}\right)^{-1}Z'u &= O_p\left(\sqrt{N}\right)O_p\left(T^{-1}\right)O_p\left(\sqrt{NT}\right)\\ &= O_p\left(T^{1/2}\right),\\ \left(\sum_{i=1}^N\sum_{t=1}^T v_{it}Z_{it}'\right)\left(Z'Z\right)^{-1}\left(\sum_{i=1}^N \overline{Z}_i\mu_i\right) &= O_p\left(\sqrt{NT}\right)O_p\left(N^{-1}T^{-1}\right)O_p\left(\sqrt{N}\right)\\ &= O_p\left(T^{-1/2}\right),\end{aligned}$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

$$\left(\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it} Z_{it}'\right)\left(Z'Z\right)^{-1} Z'u = O_p\left(\sqrt{NT}\right) O_p\left(N^{-1}T^{-1}\right) O_p\left(\sqrt{NT}\right)$$
$$= O_p(1),$$
$$\left(\sum_{i=1}^{N} \eta_i \overline{Z}_i'\right)\left(\frac{Z'Z}{T^2}\right)^{-1}\left(\sum_{i=1}^{N} \overline{Z}_i \mu_i\right) = O_p\left(\sqrt{N}\right) O_p(1)\, O_p\left(\sqrt{N}\right) = O_p(T).$$

Now, consider the case where $\delta_1 + \delta_2 > 1/2$. In this case,

$$\frac{Y_2' P_Z Y_2}{N} = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \eta_i \overline{Z}_i'\right)\left(\frac{Z'Z}{T^2}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \overline{Z}_i \eta_i'\right) + o_p(1) \tag{23}$$

and

$$\frac{Y_2' P_Z\left[(I_N \otimes \iota_T)\mu + u\right]}{N} = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \eta_i \overline{Z}_i'\right)\left(\frac{Z'Z}{T^2}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \overline{Z}_i \mu_i\right)$$
$$+ O_p\left(\max\left\{T^{-1/2}, T^{(1/2-\delta_1-\delta_2)}\right\}\right), \tag{24}$$

so that, using part (b) of Assumption 2 and applying parts (c) and (d) of Lemma S1, we obtain

$$\begin{aligned}
\widehat{\beta}_{2SLS} - \beta^0 &= \left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \eta_i \overline{Z}_i'\right)\left(\frac{Z'Z}{T^2}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \overline{Z}_i \eta_i'\right)\right]^{-1} \\
&\quad \times \left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \eta_i \overline{Z}_i'\right)\left(\frac{Z'Z}{T^2}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \overline{Z}_i \mu_i\right)\right] + o_p(1) \\
&\Longrightarrow \left(c^{-1}X^{*\prime}M^{-1}X^*\right)^{-1}\left(c^{-1}X^{*\prime}M^{-1}R^*\right) \\
&\equiv \left(X'MX\right)^{-1}\left(X'MR\right),
\end{aligned}$$

where $vec(X^*) \sim N(0, \Sigma_{\eta\eta} \otimes M_P)$, $R^* \sim N(0, \sigma_\mu^2 M_P)$, and $M^{-1}R^* \sim N\left(0, \sigma_\mu^2 M^{-1} M_P M^{-1}\right)$ and where $X = M^{-1}X^*$ and M are as defined in the statement of the theorem. Now, conditional on X,

$$\left(X'MX\right)^{-1}\left(X'MR\right) \sim N\left(0, \sigma_\mu^2\left(X'MX\right)^{-1} X' M_P X \left(X'MX\right)^{-1}\right),$$

where we have make use of the fact that X and R are independent under Assumption 5. The desired result for part (a) then follows from integrating with respect to the marginal distribution of X.

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Next, consider part (b) of the theorem, i.e., the case where $\delta_1 + \delta_2 = 1/2$. In this case,

$$\begin{aligned}
\widehat{\beta}_{2SLS} - \beta^0 &= \left[\frac{D^{0\prime}Z'ZD^0}{N^{1+2\delta_1}T^{2\delta_2}} + \frac{T^{1-\delta_2}}{N^{1/2+\delta_1}}D^{0\prime}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_i\eta_i'\right)\right. \\
&\quad + \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta_i\overline{Z}_i'\right)D^0\frac{T^{1-\delta_2}}{N^{1/2+\delta_1}} \\
&\quad \left. + \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta_i\overline{Z}_i'\right)\left(\frac{Z'Z}{T^2}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_i\eta_i'\right)\right]^{-1} \\
&\quad \times \left[\frac{T^{1-\delta_2}}{N^{1/2+\delta_1}}D^{0\prime}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_i\mu_i\right) + \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta_i\overline{Z}_i'\right)\left(\frac{Z'Z}{T^2}\right)^{-1}\right. \\
&\quad \left. \times \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_i\mu_i\right)\right] + o_p(1) \\
&\Longrightarrow \left[c^{-2\delta_1}D^{0\prime}MD^0 + c^{-1/2-\delta_1}D^{0\prime}MX + c^{-1/2-\delta_1}X'MD^0 + c^{-1}X'MX\right]^{-1} \\
&\quad \times \left[c^{-1/2-\delta_1}D^{0\prime}MR + c^{-1}X'MR\right] \\
&\equiv \left[\left(c^{1/2-\delta_1}D^0 + X\right)'M\left(c^{1/2-\delta_1}D^0 + X\right)\right]^{-1}\left(c^{1/2-\delta_1}D^0 + X\right)'MR \\
&\equiv \left[A(X)'MA(X)\right]^{-1}A(X)'MR,
\end{aligned}$$

where $A(X) = c^{1/2-\delta_1}D^0 + M^{-1}X$, as defined in the body of the theorem. Conditional on X, we have

$$\begin{aligned}
&\left[A(X)'MA(X)\right]^{-1}A(X)'MR \\
&\sim N\left(0, \sigma_\mu^2\left[A(X)'MA(X)\right]^{-1}A(X)'M_PA(X)\left[A(X)'MA(X)\right]^{-1}\right),
\end{aligned}$$

so that again the desired result follows from integrating with respect to the marginal distribution of X.

Finally, consider the case where $0 \le \delta_1 + \delta_2 < 1/2$. In this case,

$$\begin{aligned}
&T^{1/2-\delta_1-\delta_2}\left(\widehat{\beta}_{2SLS} - \beta^0\right) \\
&\quad = \left[\frac{\Pi'Z'Z\Pi}{T^{2(1-\delta_1-\delta_2)}}\right]^{-1}\left[\frac{1}{T^{3/2-\delta_1-\delta_2}}\Pi'\left(\sum_{i=1}^{N}T\overline{Z}_i\mu_i\right)\right] + o_p(1) \\
&\quad = \left[\frac{T^{2\delta_1-1}}{N^{2\delta_1-1}}\frac{D^{0\prime}Z'ZD^0}{NT}\right]^{-1}\left[\frac{N^{1/2-\delta_1}}{T^{1/2-\delta_1}}D^{0\prime}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_i\mu_i\right)\right] + o_p(1) \\
&\Longrightarrow \left[c^{-2\delta_1+1}D^{0\prime}MD^0\right]^{-1}c^{1/2-\delta_1}D^{0\prime}MR + o_p(1) \\
&\quad \equiv N\left(0, c^{2\delta_1-1}\sigma_\mu^2\left[D^{0\prime}MD^0\right]^{-1}D^{0\prime}M_PD^0\left[D^{0\prime}MD^0\right]^{-1}\right),
\end{aligned}$$

as desired. ■

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Proof of Theorem 4.2.
To proceed, write

$$\widehat{\beta}_{BG} - \beta^0 = \left(Y_2' \left[I_N \otimes P_{\iota_T}\right] Y_2\right)^{-1} Y_2' \left[\left(I_N \otimes \iota_T\right)\mu + \left(I_N \otimes P_{\iota_T}\right) u\right]. \tag{26}$$

Focusing on the denominator of (26) first, note that

$$\begin{aligned} Y_2' \left[I_N \otimes P_{\iota_T}\right] Y_2 &= \left(Z\Pi + \left[I_N \otimes \iota_T\right]\eta + V\right)' \left[I_N \otimes P_{\iota_T}\right] \left(Z\Pi + \left[I_N \otimes \iota_T\right]\eta + V\right) \\ &= \frac{D^{0\prime} Z' \left[I_N \otimes P_{\iota_T}\right] Z D^0}{N^{2\delta_1} T^{2\delta_2}} + T\eta'\eta + V' \left[I_N \otimes P_{\iota_T}\right] V \\ &\quad + \frac{D^{0\prime} Z' \left[I_N \otimes \iota_T\right]\eta}{N^{\delta_1} T^{\delta_2}} + \frac{\eta' \left[I_N \otimes \iota_T\right] Z D^0}{N^{\delta_1} T^{\delta_2}} \\ &\quad + \frac{D^{0\prime} Z' \left[I_N \otimes P_{\iota_T}\right] V}{N^{\delta_1} T^{\delta_2}} + \frac{V' \left[I_N \otimes P_{\iota_T}\right] Z D^0}{N^{\delta_1} T^{\delta_2}} + \eta' \left[I_N \otimes \iota_T'\right] V \\ &\quad + V' \left[I_N \otimes \iota_T\right]\eta, \end{aligned}$$

where, by using Assumption 2(b) and Lemmas S1 and S3, we have that

$$\begin{aligned} \frac{D^{0\prime} Z' \left[I_N \otimes P_{\iota_T}\right] Z D^0}{N^{2\delta_1} T^{2\delta_2}} &= O_p\left(T^{2(1-\delta_1-\delta_2)}\right),\ T\eta'\eta = O_p\left(T^2\right), \\ V' \left[I_N \otimes P_{\iota_T}\right] V &= O_p(T), \\ \frac{D^{0\prime} Z' \left[I_N \otimes \iota_T\right]\eta}{N^{\delta_1} T^{\delta_2}} &= O_p\left(T^{(3/2-\delta_1-\delta_2)}\right), \\ \frac{D^{0\prime} Z' \left[I_N \otimes P_{\iota_T}\right] V}{N^{\delta_1} T^{\delta_2}} &= O_p\left(T^{(1-\delta_1-\delta_2)}\right), \\ V' \left[I_N \otimes \iota_T\right]\eta &= O_p(T). \end{aligned}$$

Similarly, for the numerator of (26), note that

$$\begin{aligned} &Y_2' \left[\left(I_N \otimes \iota_T\right)\mu + \left(I_N \otimes P_{\iota_T}\right) u\right] \\ &\quad = \left(Z\Pi + \left[I_N \otimes \iota_T\right]\eta + V\right)' \left[\left(I_N \otimes \iota_T\right)\mu + \left(I_N \otimes P_{\iota_T}\right) u\right] \\ &\quad = \frac{D^{0\prime} Z' \left[I_N \otimes \iota_T\right]\mu}{N^{\delta_1} T^{\delta_2}} + \frac{D^{0\prime} Z' \left[I_N \otimes P_{\iota_T}\right] u}{N^{\delta_1} T^{\delta_2}} + T\eta'\mu + \eta' \left[I_N \otimes \iota_T'\right] u \\ &\qquad + V' \left[I_N \otimes \iota_T\right]\mu + V' \left[I_N \otimes P_{\iota_T}\right] u, \end{aligned}$$

where, by using Assumptions 2(b) and 5 and Lemmas S1–S3, we have that

$$\begin{aligned} \frac{D^{0\prime} Z' \left[I_N \otimes \iota_T\right]\mu}{N^{\delta_1} T^{\delta_2}} &= O_p\left(T^{(3/2-\delta_1-\delta_2)}\right), \\ \frac{D^{0\prime} Z' \left[I_N \otimes P_{\iota_T}\right] u}{N^{\delta_1} T^{\delta_2}} &= O_p\left(T^{(1-\delta_1-\delta_2)}\right), \\ T\eta'\mu &= O_p\left(T^{3/2}\right), \\ \eta' \left[I_N \otimes \iota_T'\right] u = O_p(T),\ V' \left[I_N \otimes \iota_T\right]\mu &= O_p(T), \\ V' \left[I_N \otimes P_{\iota_T}\right] u &= O_p(T). \end{aligned}$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

To show part (a), we consider the case where $\delta_1 + \delta_2 > 0$. In this case, in light of Lemma S7, we have

$$\begin{aligned}\sqrt{N}\left(\widehat{\beta}_{BG} - \beta^0\right) &= \left[\frac{T}{N}\frac{\eta'\eta}{N}\right]^{-1}\left[\frac{T}{N}\frac{\eta'\mu}{\sqrt{N}}\right] + o_p(1) \\ &\Longrightarrow \left[c^{-1}\Sigma_{\eta\eta}\right]^{-1} N\left(0, c^{-2}\sigma_\mu^2\Sigma_{\eta\eta}\right) \equiv N\left(0, \sigma_\mu^2\Sigma_{\eta\eta}^{-1}\right),\end{aligned}$$

as desired.

On the other hand, if $\delta_1 + \delta_2 = 0$, we have in light of Assumptions 1, 2, and 4 and Lemmas S1, S2, and S7

$$\begin{aligned}\sqrt{N}\left(\widehat{\beta}_{BG} - \beta^0\right) &= \left[\frac{T}{N^{1+2\delta_1}T^{2\delta_2}}\frac{D^{0\prime}Z'\left[I_N \otimes P_{\iota_T}\right]ZD^0}{NT} + \frac{T}{N}\frac{\eta'\eta}{N}\right]^{-1} \\ &\quad \times \left[\frac{T}{N^{1+\delta_1}T^{\delta_2}}\frac{D^{0\prime}Z'\left[I_N \otimes \iota_T\right]\mu}{T\sqrt{N}} + \frac{T}{N}\frac{\eta'\mu}{\sqrt{N}}\right] + o_p(1) \\ &= \left[\frac{T^{1+2\delta_1}}{N^{1+2\delta_1}}\frac{D^{0\prime}Z'\left[I_N \otimes P_{\iota_T}\right]ZD^0}{NT} + \frac{T}{N}\frac{\eta'\eta}{N}\right]^{-1} \\ &\quad \times \left[\frac{T^{1+\delta_1}}{N^{1+\delta_1}}\frac{D^{0\prime}Z'\left[I_N \otimes \iota_T\right]\mu}{T\sqrt{N}} + \frac{T}{N}\frac{\eta'\mu}{\sqrt{N}}\right] + o_p(1) \\ &\Longrightarrow \left[c^{-(1+2\delta_1)}M_P + c^{-1}\Sigma_{\eta\eta}\right]^{-1} N\left(0, \sigma_\mu^2\left[c^{-2(1+\delta_1)}M_P + c^{-2}\Sigma_{\eta\eta}\right]\right) \\ &\equiv N\left(0, \sigma_\mu^2\left[c^{-2\delta_1}D^{0\prime}M_P D^0 + \Sigma_{\eta\eta}\right]^{-1}\right),\end{aligned}$$

as desired. ■

Proof of Lemma 4.4. To proceed, write

$$\begin{aligned}\overline{\beta} &= \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right]Y_2}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right]Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right]Z\right)^{-1}Z'\left[I_N \otimes Q_{\iota_T}\right]Y_2\right]^{-1} \\ &\quad \times \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right]y_1}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right]Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right]Z\right)^{-1}Z'\left[I_N \otimes Q_{\iota_T}\right]y_1\right] \\ &= \beta^0 + \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right]Y_2}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right]Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right]Z\right)^{-1}Z'\left[I_N \otimes Q_{\iota_T}\right]Y_2\right]^{-1} \\ &\quad \times \frac{Y_2'[I_N \otimes \iota_T]\mu}{T} + \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right]Y_2}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right]Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right]Z\right)^{-1}\right. \\ &\qquad \left. \times Z'\left[I_N \otimes Q_{\iota_T}\right]Y_2\right]^{-1}\left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right]u}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right]\right. \\ &\qquad \left. \times Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right]Z\right)^{-1}Z'\left[I_N \otimes Q_{\iota_T}\right]u\right].\end{aligned}$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

First, consider the situation where $\delta_1 + \delta_2 \geq 1$. In this case, making use of Lemmas S7, S12–S14, and Assumption 2, we have that

$$\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2}{T} = O_p(N), \qquad \frac{Y_2'\left[I_N \otimes \iota_T\right] \mu}{T} = O_p\left(\sqrt{N}\right),$$
$$\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] u}{T} = O_p(1),$$
$$Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z \left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2 = O_p(1),$$
$$Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z \left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] u = O_p(1),$$

so that, in this case,

$$\begin{aligned}\overline{\beta} - \beta^0 = {} & \frac{1}{\sqrt{N}} \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2}{NT}\left(1 + O_p\left(N^{-1}\right)\right)\right]^{-1} \frac{Y_2'\left[I_N \otimes \iota_T\right] \mu}{\sqrt{N}T} \\ & + \frac{1}{N}\left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2}{NT}\left(1 + O_p\left(N^{-1}\right)\right)\right]^{-1} \\ & \times \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] u}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z \left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} \right. \\ & \left. \times Z'\left[I_N \otimes Q_{\iota_T}\right] u\right] = O_p\left(N^{-1/2}\right). \end{aligned} \tag{26}$$

Now, consider the case where $1/2 < \delta_1 + \delta_2 < 1$. Here,

$$\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2}{T} = O_p(N), \qquad \frac{Y_2'\left[I_N \otimes \iota_T\right] \mu}{T} = O_p\left(\sqrt{N}\right),$$
$$\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] u}{T} = O_p(1),$$
$$\begin{aligned} Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z \left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2 &= O_p\left(N^{1-2\delta_1} T^{1-2\delta_2}\right) \\ &= O_p\left(T^{2[1-(\delta_1+\delta_2)]}\right), \\ Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z \left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] u &= O_p\left(N^{1/2-\delta_1} T^{1/2-\delta_2}\right) \\ &= O_p\left(T^{1-(\delta_1+\delta_2)}\right), \end{aligned}$$

so that, in this case,

$$\begin{aligned}\overline{\beta} - \beta^0 = {} & \frac{1}{\sqrt{N}} \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2}{NT}\left(1 + O_p\left(N^{-2(\delta_1+\delta_2)+1}\right)\right)\right]^{-1} \frac{Y_2'\left[I_N \otimes \iota_T\right] \mu}{T\sqrt{N}} \\ & + \frac{1}{N^{\delta_1+\delta_2}}\left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2}{NT}\left(1 + O_p\left(N^{-2(\delta_1+\delta_2)+1}\right)\right)\right]^{-1} \\ & \times \left[\frac{Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z \left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] u}{N^{1-(\delta_1+\delta_2)}} \right. \\ & \left. \times \left(1 + O_p\left(N^{-1+(\delta_1+\delta_2)}\right)\right)\right] = O_p\left(N^{-1/2}\right). \end{aligned} \tag{27}$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Next, consider the case where $\delta_1 + \delta_2 = 1/2$. In this case,

$$\frac{Y_2'[I_N \otimes P_{\iota_T}]Y_2}{T} = O_p(N), \qquad \frac{Y_2'[I_N \otimes \iota_T]\mu}{T} = O_p(\sqrt{N}),$$
$$\frac{Y_2'[I_N \otimes P_{\iota_T}]u}{T} = O_p(1),$$
$$Y_2'[I_N \otimes Q_{\iota_T}]Z\left(Z'[I_N \otimes Q_{\iota_T}]Z\right)^{-1}Z'[I_N \otimes Q_{\iota_T}]Y_2$$
$$= O_p\left(N^{1-2\delta_1}T^{1-2\delta_2}\right) = O_p(N),$$
$$Y_2'[I_N \otimes Q_{\iota_T}]Z\left(Z'[I_N \otimes Q_{\iota_T}]Z\right)^{-1}Z'[I_N \otimes Q_{\iota_T}]u$$
$$= O_p\left(N^{1/2-\delta_1}T^{1/2-\delta_2}\right) = O_p(\sqrt{N}).$$

Hence,

$$\begin{aligned}\overline{\beta} - \beta^0 = \frac{1}{\sqrt{N}}\Bigg[&\frac{Y_2'[I_N \otimes P_{\iota_T}]Y_2}{NT} \\ &+ \frac{Y_2'[I_N \otimes Q_{\iota_T}]Z\left(Z'[I_N \otimes Q_{\iota_T}]Z\right)^{-1}Z'[I_N \otimes Q_{\iota_T}]Y_2}{N}\Bigg]^{-1} \\ &\times \frac{Y_2'[I_N \otimes \iota_T]\mu}{T\sqrt{N}} + \frac{1}{\sqrt{N}}\Bigg[\frac{Y_2'[I_N \otimes P_{\iota_T}]Y_2}{NT} \\ &+ \frac{Y_2'[I_N \otimes Q_{\iota_T}]Z\left(Z'[I_N \otimes Q_{\iota_T}]Z\right)^{-1}Z'[I_N \otimes Q_{\iota_T}]Y_2}{N}\Bigg]^{-1} \\ &\times \Bigg[\frac{Y_2'[I_N \otimes Q_{\iota_T}]Z\left(Z'[I_N \otimes Q_{\iota_T}]Z\right)^{-1}Z'[I_N \otimes Q_{\iota_T}]u}{\sqrt{N}} \\ &\times \left(1 + O_p\left(N^{-1/2}\right)\right)\Bigg] \\ = O_p&\left(N^{-1/2}\right). \end{aligned} \tag{28}$$

Finally, consider the case $0 \le \delta_1 + \delta_2 < 1/2$. In this case,

$$\frac{Y_2'[I_N \otimes P_{\iota_T}]Y_2}{T} = O_p(N), \qquad \frac{Y_2'[I_N \otimes \iota_T]\mu}{T} = O_p(\sqrt{N}),$$
$$\frac{Y_2'[I_N \otimes P_{\iota_T}]u}{T} = O_p(1),$$
$$Y_2'[I_N \otimes Q_{\iota_T}]Z\left(Z'[I_N \otimes Q_{\iota_T}]Z\right)^{-1}Z'[I_N \otimes Q_{\iota_T}]Y_2 = O_p\left(N^{1-2\delta_1}T^{1-2\delta_2}\right)$$
$$= O_p\left(N^{2-2(\delta_1+\delta_2)}\right),$$
$$Y_2'[I_N \otimes Q_{\iota_T}]Z\left(Z'[I_N \otimes Q_{\iota_T}]Z\right)^{-1}Z'[I_N \otimes Q_{\iota_T}]u = O_p\left(N^{1/2-\delta_1}T^{1/2-\delta_2}\right)$$
$$= O_p\left(N^{1-(\delta_1+\delta_2)}\right),$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

so that

$$\begin{aligned}\overline{\beta}-\beta^0 &= \frac{1}{N^{2-2(\delta_1+\delta_2)-1/2}}\left[\frac{Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\left(Z'\left[I_N\otimes Q_{\iota_T}\right]Z\right)^{-1}Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2}{N^{2-2(\delta_1+\delta_2)}}\right.\\ &\quad\left.\times\left(1+O_p\left(N^{-1+2(\delta_1+\delta_2)}\right)\right)\right]^{-1}\frac{Y_2'\left[I_N\otimes\iota_T\right]\mu}{T\sqrt{N}}\\ &\quad+\frac{1}{N^{1-(\delta_1+\delta_2)}}\left[\frac{Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\left(Z'\left[I_N\otimes Q_{\iota_T}\right]Z\right)^{-1}Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2}{N^{2-2(\delta_1+\delta_2)}}\right.\\ &\quad\left.\times\left(1+O_p\left(N^{-1+2(\delta_1+\delta_2)}\right)\right)\right]^{-1}\\ &\quad\times\left[\frac{Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\left(Z'\left[I_N\otimes Q_{\iota_T}\right]Z\right)^{-1}Z'\left[I_N\otimes Q_{\iota_T}\right]u}{N^{1-(\delta_1+\delta_2)}}\right]\\ &\quad\times\left(1+O_p\left(N^{-1+(\delta_1+\delta_2)}\right)\right)\\ &= O_p\left(N^{-1+(\delta_1+\delta_2)}\right). \end{aligned}\tag{29}$$

Together, (26)-(29) imply that $\overline{\beta}\xrightarrow{p}\beta^0$ for all $\delta_1\geq 0$ and $\delta_2\geq 0$, as desired. ■

Proof of Theorem 4.5. To show part (a), write

$$\begin{aligned}&\widehat{\Xi}_{NT}^{1/2}\left(\widehat{\beta}_{GMM}-\beta^0\right)\\ &\quad=\left\{\left[\widehat{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\right.\\ &\qquad\left.\times\left[\widehat{\sigma}_\mu^{-2}\frac{Y_2'[I_N\otimes\iota_T]\mu}{T}+\widehat{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\right\}.\end{aligned}$$

Now, from Lemmas S5, S7, and S12–S14 and the continuous mapping theorem, we have that for $\delta_1+\delta_2>1/2$

$$\begin{aligned}&\widehat{\Xi}_{NT}^{1/2}\left(\widehat{\beta}_{GMM}-\beta^0\right)\\ &\quad=\left[\widehat{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\ &\qquad\times\left[\widehat{\sigma}_\mu^{-2}\frac{Y_2'[I_N\otimes\iota_T]\mu}{T}+\widehat{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\\ &\quad=\left[\widehat{\sigma}_2^{-2}\eta'\eta+O_p(1)+O_p\left(\max\left\{1,T^{2(1-\delta_1-\delta_2)}\right\}\right)\right]^{-1/2}\\ &\qquad\times\left[\widehat{\sigma}_\mu^{-2}\eta'\mu+O_p(1)+O_p\left(\max\left\{1,T^{1-\delta_1-\delta_2}\right\}\right)\right]\\ &\quad=\left[\widehat{\sigma}_\mu^{-2}\eta'\eta/N\right]^{-1/2}\left(1+O_p\left(N^{-1}\right)+O_p\left(T^{1-2(\delta_1+\delta_2)}\right)\right)\end{aligned}$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

$$\times \left[\widehat{\sigma}_{\mu}^{-2}\eta'\mu/\sqrt{N}\right]\left(1+O_p\left(\max\left\{T^{-1/2},T^{1/2-(\delta_1+\delta_2)}\right\}\right)\right)$$
$$=\widehat{\sigma}_{\mu}^{-1}\left[\eta'\eta/N\right]^{-1/2}\left[\eta'\mu/\sqrt{N}\right]\left(1+O_p\left(\frac{1}{\min\left\{\sqrt{T},T^{(\delta_1+\delta_2)-1/2}\right\}}\right)\right)$$
$$\Longrightarrow N\left(0,I_G\right), \tag{30}$$

as $N,T\to\infty$ such that $N/T\to c\neq 0$.

To show part (b), note that for the case where $\delta_1+\delta_2=1/2$, we have in lieu of (30)

$$\widehat{\Xi}_{NT}^{1/2}\left(\widehat{\beta}_{GMM}-\beta^0\right)$$
$$=\left[\widehat{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[\widehat{\sigma}_{\mu}^{-2}\overline{Y}_2'\mu+\widehat{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]$$
$$=\left[\widehat{\sigma}_{\mu}^{-2}\eta'\eta+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2+O_p\left(1\right)\right]^{-1/2}$$
$$\times\left[\widehat{\sigma}_{\mu}^{-2}\eta'\mu+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u+O_p\left(1\right)\right]$$
$$=\left[\widehat{\sigma}_{\mu}^{-2}\eta'\eta+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[\widehat{\sigma}_{\mu}^{-2}\eta'\mu+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\left(1+O_p\left(N^{-1/2}\right)\right)$$
$$\Longrightarrow N\left(0,I_G\right),$$

as $N,T\to\infty$ such that $N/T\to c\neq 0$.

Finally, for part (c), note that, for $\delta_1+\delta_2<1/2$, we can apply Lemmas S5, S7, and S12–S14, and the continuous mapping theorem to obtain

$$\widehat{\Xi}_{NT}^{1/2}\left(\widehat{\beta}_{GMM}-\beta^0\right)$$
$$=\left[\widehat{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[\widehat{\sigma}_{\mu}^{-2}\overline{Y}_2'\mu+\widehat{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]$$
$$=\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2+O_p\left(N\right)\right]^{-1/2}$$
$$\times\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u+O_p\left(\sqrt{N}\right)\right]$$
$$=\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\left(1+O_p\left(N^{-1/2+(\delta_1+\delta_2)}\right)\right)$$
$$\Longrightarrow N\left(0,I_G\right),$$

as $N,T\to\infty$ such that $N/T\to c\neq 0$. ■

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Proof of Theorem 5.1.
To show part (a), note first that by part (a) of Lemma S13, we have for $\delta_1+\delta_2>0$

$$\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{NT}=\frac{\eta'\eta}{N}+O_p\left(\max\left\{N^{-\delta_1}T^{-\delta_2},N^{-1}\right\}\right)\xrightarrow{p}\Sigma_{\eta\eta}.$$

Moreover, by part (b) of Lemma S13 and part (b) of Lemma S14, we have that

$$\begin{aligned}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{NT}&=\left(\frac{1}{N}\sum_{i=1}^{N}\eta_i\overline{u}_i+\frac{1}{N}\sum_{i=1}^{N}\overline{v}_i\overline{u}_i\right)\left(1+O_p\left(\frac{1}{N^{\delta_1}T^{\delta_2}}\right)\right)\\&=O_p\left(N^{-1}\right),\\\frac{Y_2'\left[I_N\otimes\iota_T\right]\mu}{NT}&=\frac{1}{N}\sum_{i=1}^{N}\eta_i\mu_i+O_p\left(\max\left\{\frac{1}{N},\frac{1}{N^{1/2+\delta_1}T^{\delta_2}}\right\}\right)\xrightarrow{p}\sigma_{\eta\mu},\end{aligned}$$

for $\delta_1+\delta_2>0$. It follows by Slutsky's theorem that

$$\begin{aligned}\widehat{\beta}_{BG}-\beta^0&=\left(\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{NT}\right)^{-1}\left[\frac{Y_2'\left(I_N\otimes\iota_T\right)\mu}{NT}+\frac{Y_2'\left(I_N\otimes P_{\iota_T}\right)u}{NT}\right]\\&\xrightarrow{p}\Sigma_{\eta\eta}^{-1}\sigma_{\eta\mu},\end{aligned}$$

as desired.

For part (b), we apply parts (a) and (b) of Lemma S13 and part (b) of Lemma S14 for the case $\delta_1+\delta_2=0$ to obtain

$$\begin{aligned}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{NT}&=\frac{\eta'\eta}{N}+\frac{D^0Z'\left[I_N\otimes P_{\iota_T}\right]ZD^0}{NT}+O_p\left(\frac{1}{\sqrt{N}}\right)\\&\xrightarrow{p}\Sigma_{\eta\eta}+D^0M_PD^0,\\\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{NT}&=\frac{1}{N}\sum_{i=1}^{N}\eta_i\overline{u}_i+\frac{1}{N}\sum_{i=1}^{N}\overline{v}_i\overline{u}_i+\frac{1}{N}\sum_{i=1}^{N}D^0Z'\overline{u}_i=O_p\left(N^{-1}\right),\\\frac{Y_2'\left[I_N\otimes\iota_T\right]\mu}{NT}&=\frac{1}{N}\sum_{i=1}^{N}\eta_i\mu_i+O_p\left(\max\left\{\frac{1}{N},\frac{1}{N^{1/2+\delta_1}T^{\delta_2}}\right\}\right)\xrightarrow{p}\sigma_{\eta\mu}.\end{aligned}$$

Hence, the Slutsky's theorem implies that

$$\begin{aligned}\widehat{\beta}_{BG}-\beta^0&=\left(\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{NT}\right)^{-1}\left[\frac{Y_2'\left(I_N\otimes\iota_T\right)\mu}{NT}+\frac{Y_2'\left(I_N\otimes P_{\iota_T}\right)u}{NT}\right]\\&\xrightarrow{p}\left(D^0M_PD^0+\Sigma_{\eta\eta}\right)^{-1}\sigma_{\eta\mu},\end{aligned}$$

which is the result desired. ■

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Proof of Theorem 5.2.
To proceed, write

$$\begin{aligned}\overline{\beta} &= \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2\right]^{-1} \\ &\quad \times \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] y_1}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] y_1\right] \\ &= \beta^0 + \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2\right]^{-1} \\ &\quad \times \frac{Y_2'[I_N \otimes \iota_T]\mu}{T} + \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1}\right. \\ &\qquad \left. \times Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2\right]^{-1} \\ &\quad \times \left[\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] u}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] u\right].\end{aligned}$$

Given that $0 \le \delta_1 + \delta_2 < 1/2$, we have from Lemmas S12 and S13, part (b) of Lemma S14, and Assumption 2(b) that

$$\begin{aligned}&\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2}{T} = O_p(N), \qquad \frac{Y_2'\left[I_N \otimes \iota_T\right] \mu}{T} = O_p(N), \\ &\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] u}{T} = O_p(1), \\ &\frac{Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2}{N^{2-2(\delta_1+\delta_2)}} \xrightarrow{p} D^{0\prime} M_Q D^0, \\ &Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] u = O_p\left(N^{1/2-\delta_1} T^{1/2-\delta_2}\right) \\ &\qquad\qquad = O_p\left(N^{1-(\delta_1+\delta_2)}\right),\end{aligned}$$

so that

$$\begin{aligned}\overline{\beta} - \beta^0 &= \frac{1}{N^{2-2(\delta_1+\delta_2)-1/2}} \\ &\quad \times \left[\frac{Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2}{N^{2-2(\delta_1+\delta_2)}}\right. \\ &\quad \left. \times \left(1 + O_p\left(N^{-1+2(\delta_1+\delta_2)}\right)\right)\right]^{-1} \\ &\quad \times \frac{Y_2'\left[I_N \otimes \iota_T\right] \mu}{T\sqrt{N}} + \frac{1}{N^{1-(\delta_1+\delta_2)}} \\ &\quad \times \left[\frac{Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z\left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2}{N^{2-2(\delta_1+\delta_2)}}\right.\end{aligned}$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

$$\times \left(1 + O_p\left(N^{-1+2(\delta_1+\delta_2)}\right)\right)\Big]^{-1}$$
$$\times \left[\frac{Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z \left(Z'\left[I_N \otimes Q_{\iota_T}\right] Z\right)^{-1} Z'\left[I_N \otimes Q_{\iota_T}\right] u}{N^{1-(\delta_1+\delta_2)}}\right]$$
$$\times \left(1 + O_p\left(N^{-1+(\delta_1+\delta_2)}\right)\right) = O_p\left(N^{-1+(\delta_1+\delta_2)}\right) \xrightarrow{p} 0,$$

as desired. ■

Proof of Theorem 5.3.

Note that, under Assumption 8, we can apply Lemma S6 to obtain

$$NT\widehat{W}_2 = \left[\frac{1}{N}\sum_{i=1}^{N} \widehat{J}_{i,T}\right] \xrightarrow{p} \overline{J}^{-1}, \text{ as } N, T \to \infty \text{ such that } \frac{N}{T} \to c.$$

To show part (a), note that for $\delta_1 + \delta_2 = 0$, we have

$$\widehat{\Xi}_{NT}^{1/2}\left(\widehat{\beta}_{GMM} - \beta^0\right)$$
$$= \left\{\left[\widehat{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] Y_2}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z\widehat{W}_2 Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2\right]^{-1/2}\right.$$
$$\times \left[\widehat{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N \otimes \iota_T\right]\mu}{T} + \widehat{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N \otimes P_{\iota_T}\right] u}{T} + Y_2'\left[I_N \otimes Q_{\iota_T}\right]\right.$$
$$\left.\left. \times Z\widehat{W}_2 Z'\left[I_N \otimes Q_{\iota_T}\right] u\right]\right\}$$
$$= \left(\frac{N}{T}\right)^{\frac{1}{2}+\delta_1}\left[\frac{Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z}{N^{1-\delta_1}T^{1-\delta_2}}\overline{J}^{-1}\frac{Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2}{N^{1-\delta_1}T^{1-\delta_2}}\right]^{-1/2}\widehat{\sigma}_\mu^{-2}\frac{\eta'\mu}{N}\left(1 + o_p(1)\right)$$
$$+ \left[\frac{Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z}{N^{1-\delta_1}T^{1-\delta_2}}\overline{J}^{-1}\frac{Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2}{N^{1-\delta_1}T^{1-\delta_2}}\right]^{-1/2}$$
$$\times \left[\frac{Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z}{N^{1-\delta_1}T^{1-\delta_2}}\overline{J}^{-1/2}\left(\overline{J}^{-1/2}\frac{Z'\left[I_N \otimes Q_{\iota_T}\right] u}{T}\right)\right]\left(1 + o_p(1)\right).$$

Since

$$\left[\frac{Y_2'\left[I_N \otimes Q_{\iota_T}\right] Z}{N^{1-\delta_1}T^{1-\delta_2}}\overline{J}^{-1}\frac{Z'\left[I_N \otimes Q_{\iota_T}\right] Y_2}{N^{1-\delta_1}T^{1-\delta_2}}\right]^{-\frac{1}{2}}\left(\frac{N}{T}\right)^{\frac{1}{2}+\delta_1}\frac{\eta'\mu}{\widehat{\sigma}_\mu^2 N}$$
$$\xrightarrow{p} \left(D^{0\prime}M_Q\overline{J}^{-1}M_Q D^0\right)^{-\frac{1}{2}}\frac{c^{\frac{1}{2}+\delta_1}\sigma_{\eta\mu}}{\sigma_\mu^2} = \vartheta_1$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

and

$$\left[\frac{Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z}{N^{1-\delta_1}T^{1-\delta_2}}\overline{J}^{-1}\frac{Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2}{N^{1-\delta_1}T^{1-\delta_2}}\right]^{-1/2}\left[\frac{Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z}{N^{1-\delta_1}T^{1-\delta_2}}\overline{J}^{-1/2}\right.$$
$$\left.\times\left(\overline{J}^{-1/2}\frac{Z'\left[I_N\otimes Q_{\iota_T}\right]u}{T}\right)\right]\Longrightarrow\left(D^{0\prime}M_Q\overline{J}^{-1}M_QD^0\right)^{-1/2}D^{0\prime}M_Q\overline{J}^{-1/2}N\left(0,I_K\right)$$
$$\equiv N\left(0,I_G\right),$$

we deduce by the Cramér Convergence Theorem that

$$\widehat{\Xi}_{NT}^{1/2}\left(\widehat{\beta}_{GMM}-\beta^0\right)\Longrightarrow N\left(\vartheta_1,I_G\right),$$

as desired.

To show part (b), note that, for $0<\delta_1+\delta_2<1/2$, we apply Lemmas S6 and S13 and part (b) of Lemma S14 to obtain

$$T^{1-2(\delta_1+\delta_2)}\left(\widehat{\beta}_{GMM}-\beta^0\right)$$
$$=T^{1-2(\delta_1+\delta_2)}\left\{\left[\widehat{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1}\right.$$
$$\left.\times\left[\widehat{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes\iota_T\right]\mu}{T}+\widehat{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widehat{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\right\}$$
$$=\left(\frac{N}{T}\right)^{2\delta_1}\left[\frac{Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z}{N^{1-\delta_1}T^{1-\delta_2}}\overline{J}^{-1}\frac{Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2}{N^{1-\delta_1}T^{1-\delta_2}}\right]^{-1}\widehat{\sigma}_\mu^{-2}\frac{\eta'\mu}{N}\left(1+o_p\left(1\right)\right)$$
$$+O_p\left(T^{-(\delta_1+\delta_2)}\right).$$

It follows by Lemmas S12 and part (b) of Lemma S16 that

$$T^{1-2(\delta_1+\delta_2)}\left(\widehat{\beta}_{GMM}-\beta^0\right)\xrightarrow{p}c^{2\delta_1}\left(D^{0\prime}M_Q\overline{J}^{-1}M_QD^0\right)^{-1/2}\frac{\sigma_{\eta\mu}}{\sigma_\mu^2}=\vartheta_2.\qquad\blacksquare$$

Proof of Theorem 5.5.

We will only prove the results for $\widetilde{\beta}_C$. The proof for $\widetilde{\beta}_{SBIC}$ is virtually identical to that for $\widetilde{\beta}_C$ with the minor modification that we apply Lemma S24 in lieu of Lemma S23.

To show part (a), write

$$\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_C-\beta^0\right)$$
$$=\left\{\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\right.$$
$$\left.\times\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes\iota_T\right]\mu}{T}+\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\right\}$$
$$-\left[1-P\left(\chi>\mathbb{T}_N-G\ln N\right)\right]\left\{\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{1/2}\right.$$
$$\left.\times\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1}\left[N\widetilde{\sigma}_{\eta\mu}/\widetilde{\sigma}_\mu^2\right]\right\}.\qquad(31)$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

Now, similar to the proof of part (a) of Theorem 4.5, we can apply Lemma S5, Lemmas S12–S13, part (a) of Lemma S14 to obtain, for $\delta_1+\delta_2>1/2$, the asymptotic representation

$$\begin{aligned}
&\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\
&\quad\times\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes \iota_T\right]\mu}{T}+\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\\
&\quad=\widetilde{\sigma}_\mu^{-1}\left[\eta'\eta/N\right]^{-1/2}\left[\eta'\mu/\sqrt{N}\right]\left(1+O_p\left(\frac{1}{\min\left\{\sqrt{N},T^{(\delta_1+\delta_2)-1/2}\right\}}\right)\right). \qquad \textbf{(32)}
\end{aligned}$$

Note that the main difference between (32) and expression (30) above is that in the latter case the weighting matrix $\widehat{W}_2$ is constructed using residuals $\widehat{u}_{it}=\left(y_{1,it}-\overline{y}_{1,i}\right)-\left(Y_{2,it}-\overline{Y}_{2,i}\right)'\overline{\beta}$ based on the preliminary estimator $\overline{\beta}$, whereas the former uses a weighting matrix $\widetilde{W}_2$ based on residuals constructed from the GMM estimator, i.e., $\widetilde{u}_{it}=\left(y_{1,it}-\overline{y}_{1,i}\right)-\left(Y_{2,it}-\overline{Y}_{2,i}\right)'\widehat{\beta}_{GMM}$. Since both estimators are consistent under the assumptions of this theorem, it is straightforward to show that $\widetilde{W}_2-\widehat{W}_2\xrightarrow{p}0$, and so we obtain similar asymptotic behavior in the two cases.

Moreover, also by Lemmas S5, S12–S13, and part (a) of Lemma S23

$$\begin{aligned}
&\left[1-P\left(\chi>\mathbb{T}_N-G\ln N\right)\right]\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{1/2}\\
&\quad\times\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1}\left[N\widetilde{\sigma}_{\eta\mu}/\widetilde{\sigma}_\mu^2\right]\\
&\quad=\left[1-P\left(\chi>\mathbb{T}_N-G\ln N\right)\right]\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\
&\qquad\times\left[N\widetilde{\sigma}_{\eta\mu}/\widetilde{\sigma}_\mu^2\right]\\
&\quad=o_p(1)\left[\widetilde{\sigma}_\mu^{-2}\eta'\eta+O_p(1)+O_p\left(T^{2(1-\delta_1-\delta_2)}\right)\right]^{-1/2}\\
&\qquad\times\left[\widetilde{\sigma}_\mu^{-2}\eta'\mu+\widetilde{\sigma}_\mu^{-2}\eta'\eta\left(\beta^0-\widehat{\beta}_{GMM}\right)\right]\left(1+O_p\left(N^{-1/2}\right)\right)\\
&\quad=o_p(1)\,O_p\left(N^{-1/2}\right)O_p\left(\sqrt{N}\right)=o_p(1). \qquad \textbf{(33)}
\end{aligned}$$

It follows from (32), (33), and the continuous mapping theorem that

$$\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_C-\beta^0\right)=\widetilde{\sigma}_\mu^{-1}\left[\eta'\eta/N\right]^{-1/2}\left[\eta'\mu/\sqrt{N}\right]\left(1+o_p(1)\right)\Longrightarrow N\left(0,I_G\right),$$

as $N,T\to\infty$ such that $N/T\to c\neq 0$.

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

For part (b), note that for the case where $\delta_1+\delta_2=1/2$, we have in lieu of (32) and (33)

$$\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[\widetilde{\sigma}_\mu^{-2}\overline{Y}_2'\mu+\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]$$
$$=\left[\widetilde{\sigma}_\mu^{-2}\eta'\eta+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[\widetilde{\sigma}_\mu^{-2}\eta'\mu+Y_2'[I_N\otimes Q_{\iota_T}]Z\widetilde{W}_2Z'[I_N\otimes Q_{\iota_T}]u\right]\left(1+O_p\left(T^{-1/2}\right)\right) \tag{34}$$

and

$$\left[1-P\left(\chi>\mathbb{T}_N-G\ln N\right)\right]\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{1/2}$$
$$\times\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1}\left[N\widetilde{\sigma}_{\eta\mu}/\widetilde{\sigma}_\mu^2\right]$$
$$=\left[1-P\left(\chi>\mathbb{T}_N-G\ln N\right)\right]\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[N\widetilde{\sigma}_{\eta\mu}/\widetilde{\sigma}_\mu^2\right]$$
$$=\left[1-P\left(\chi>\mathbb{T}_N-G\ln N\right)\right]\left[\widetilde{\sigma}_\mu^{-2}\eta'\eta+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2+O_p(1)\right]^{-1/2}$$
$$\times\left[\widetilde{\sigma}_\mu^{-2}\eta'\mu+\widetilde{\sigma}_\mu^{-2}\eta'\eta\left(\beta^0-\widehat{\beta}_{GMM}\right)\right]\left(1+O_p\left(N^{-1/2}\right)\right)$$
$$=o_p(1)\,O_p\left(N^{-1/2}\right)O_p\left(\sqrt{N}\right)=o_p(1). \tag{35}$$

It follows from (34), (35), and the continuous mapping theorem that

$$\begin{aligned}\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_C-\beta^0\right)&=\left[\widetilde{\sigma}_\mu^{-2}\eta'\eta+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\&\quad\times\left[\widetilde{\sigma}_\mu^{-2}\eta'\mu+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\\&\quad+O_p\left(T^{-1/2}\right)+o_p(1)\\&\Longrightarrow N\left(0,I_G\right),\end{aligned}$$

as $N,T\to\infty$ such that $N/T\to c\neq 0$.

For part (c), we consider the case $0\leq\delta_1+\delta_2<1/2$. Again, note that by argument similar to that given for part (c) of Theorem 4.5, we obtain by applying Lemmas S5, S7, S12–S13, and part (a) of Lemma S23

$$\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[\widetilde{\sigma}_\mu^{-2}\overline{Y}_2'\mu+\widehat{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]$$
$$=\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\left(1+O_p\left(N^{-1/2+(\delta_1+\delta_2)}\right)\right) \tag{36}$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

and

$$
\begin{aligned}
&[1-P(\chi>\mathbb{T}_N-G\ln N)]\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{1/2}\\
&\times\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1}\left[N\widetilde{\sigma}_{\eta\mu}/\widetilde{\sigma}_\mu^2\right]\\
&=[1-P(\chi>\mathbb{T}_N-G\ln N)]\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2+O_p(N)\right]^{-1/2}\\
&\quad\times\left[\widetilde{\sigma}_\mu^{-2}\eta'\mu\right]\left(1+O_p\left(N^{-(1/2-\delta_1-\delta_2)}\right)\right)\\
&=o_p(1)\,O_p\left(N^{-(1-\delta_1-\delta_2)}\right)O_p\left(\sqrt{N}\right)=o_p(1)\,O_p\left(N^{-(1/2-\delta_1-\delta_2)}\right)\\
&=o_p(1),
\end{aligned}
\tag{37}
$$

where the last equality follows in part from the fact that, in this case, we assume that $0\le\delta_1+\delta_2<1/2$. It follows from (36), (37), and the continuous mapping theorem that

$$
\begin{aligned}
\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_C-\beta^0\right) &= \left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\
&\quad\times\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]+o_p(1)\\
&\Longrightarrow N(0,I_G),
\end{aligned}
$$

as $N,T\to\infty$ such that $N/T\to c\neq 0$. ■

Proof of Theorem 5.6.

We will only prove the results for $\widetilde{\beta}_C$. The proof for $\widetilde{\beta}_{SBIC}$ is virtually identical to that for $\widetilde{\beta}_C$ with the minor modification that we apply Lemma S24 in lieu of Lemma S23.

Similar to the proof of Theorem 5.5, we can write

$$
\begin{aligned}
\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_C-\beta^0\right)=&\left\{\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\right.\\
&\times\left[\widetilde{\sigma}_\mu^{-2}\overline{Y}_2'\mu+\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]\right.\\
&\left.\left.\times\ Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\right\}\\
&-\left\{[1-P(\chi>\mathbb{T}_N-G\ln N)]\left[\widetilde{\sigma}_\mu^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}\right.\right.\\
&\left.\left.+Y_2'\left[I_N\otimes Q_{\iota_T}\right]\ Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}N\frac{\widetilde{\sigma}_{\eta\mu}}{\widetilde{\sigma}_\mu^2}\right\}.
\end{aligned}
$$

Note that, in this case, using Lemmas S6, S8, S12–S13, part (b) of Lemma S14, part (b) of Lemma S16, and part (b) of Lemma S23; we obtain

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

$$\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[\widetilde{\sigma}_{\mu}^{-2}\overline{Y}_2'\mu+\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]u}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]$$
$$=\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[\widetilde{\sigma}_{\mu}^{-2}\eta'\mu\left(1+O_p\left(T^{-(1/2+\delta_1+\delta_2)}\right)\right)\right]$$
$$+\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\left(1+O_p\left(T^{-1+2(\delta_1+\delta_2)}\right)\right)$$
$$=\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\widetilde{\sigma}_{\mu}^{-2}\eta'\mu+O_p\left(T^{-1/2}\right)+\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\left(1+O_p\left(T^{-1+2(\delta_1+\delta_2)}\right)\right) \tag{38}$$

and

$$\left[1-P\left(\chi>\mathbb{T}_N-G\ln N\right)\right]$$
$$\times\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}N\widetilde{\sigma}_{\eta\mu}/\widetilde{\sigma}_{\mu}^2$$
$$=\left[1-P\left(\chi>\mathbb{T}_N-G\ln N\right)\right]$$
$$\times\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\widetilde{\sigma}_{\mu}^{-2}\eta'\mu\left(1+O_p\left(N^{-1/2}\right)\right)$$
$$=\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\widetilde{\sigma}_{\mu}^{-2}\eta'\mu\left(1+O_p\left(N^{-1/2}\right)\right)+P\left(\chi>\mathbb{T}_N-G\ln N\right)$$
$$\times\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\widetilde{\sigma}_{\mu}^{-2}\eta'\mu\left(1+O_p\left(N^{-1/2}\right)\right)$$
$$=\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}$$
$$\times\widetilde{\sigma}_{\mu}^{-2}\eta'\mu+O_p\left(N^{-(1/2-\delta_1-\delta_2)}\right)+P\left(\chi>\mathbb{T}_N-G\ln N\right)$$

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press

$$
\begin{aligned}
&\times \left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\
&\times\widetilde{\sigma}_{\mu}^{-2}\eta'\mu\left(1+O_p\left(N^{-1/2}\right)\right)\\
=&\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\
&\times\widetilde{\sigma}_{\mu}^{-2}\eta'\mu+O_p\left(N^{-(1/2-\delta_1-\delta_2)}\right)+O_p\left(N^{-1}\right)O_p\left(N^{-1+\delta_1+\delta_2}\right)O_p\left(N\right)\\
=&\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\
&\times\widetilde{\sigma}_{\mu}^{-2}\eta'\mu+O_p\left(N^{-(1/2-\delta_1-\delta_2)}\right), \qquad (39)
\end{aligned}
$$

where the fourth inequality above follows from applying part (b) of Lemma S23. It follows from (38), (39), and the continuous mapping theorem that

$$
\begin{aligned}
\widetilde{\Xi}_{NT}^{1/2}\left(\widetilde{\beta}_C-\beta^0\right) &= \left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\
&\quad\times\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]\left(1+O_p\left(T^{-1+2(\delta_1+\delta_2)}\right)\right)\\
&\quad+\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\
&\quad\times\widetilde{\sigma}_{\mu}^{-2}\eta'\mu+O_p\left(T^{-1/2}\right)\\
&\quad-\left[\widetilde{\sigma}_{\mu}^{-2}\frac{Y_2'\left[I_N\otimes P_{\iota_T}\right]Y_2}{T}+Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\
&\quad\times\widetilde{\sigma}_{\mu}^{-2}\eta'\mu+O_p\left(N^{-(1/2-\delta_1-\delta_2)}\right)\\
&= \left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]Y_2\right]^{-1/2}\\
&\quad\times\left[Y_2'\left[I_N\otimes Q_{\iota_T}\right]Z\widetilde{W}_2Z'\left[I_N\otimes Q_{\iota_T}\right]u\right]+o_p\left(1\right)\\
&\Longrightarrow N\left(0,I_G\right),
\end{aligned}
$$

where the second equality above makes use of the fact that $0\leq\delta_1+\delta_2<1/2$ in this case. ■

https://doi.org/10.1017/S0266466613000492 Published online by Cambridge University Press